

Texto para Discussão

Série Economia

TD-E 03 / 2009

**Modelos De Fatores Latentes Generalizados Para Curvas De Juros Em
Múltiplos Mercados**

Prof. Dr. Márcio Poletti Laurini



Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Universidade de São Paulo

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
de Ribeirão Preto

Reitor da Universidade de São Paulo
Suely Vilela

Diretor da FEA-RP/USP
Rudinei Toneto Junior

Chefe do Departamento de Administração
André Lucirton Costa

Chefe do Departamento de Contabilidade
Adriana Maria Procópio de Araújo

Chefe do Departamento de Economia
Walter Belluzzo Junior

CONSELHO EDITORIAL

Comissão de Pesquisa da FEA-RP/USP

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Avenida dos Bandeirantes, 3900
14049-905 Ribeirão Preto - SP

A série TEXTO PARA DISCUSSÃO tem como objetivo divulgar: i) resultados de trabalhos em desenvolvimento na FEA-RP/USP; ii) trabalhos de pesquisadores de outras instituições considerados de relevância dadas as linhas de pesquisa da instituição. A série foi subdividida em função das principais áreas de atuação da FEA-RP/USP: Economia, Administração e Contabilidade. Veja o site da CPq na Home Page da FEA-RP: www.fearp.usp.br. Informações: e-mail: cpq@fearp.usp.br

MODELOS DE FATORES LATENTES GENERALIZADOS PARA CURVAS DE JUROS EM MÚLTIPLOS MERCADOS

MÁRCIO POLETTI LAURINI
IBMEC SÃO PAULO E IMECC-UNICAMP
LUIZ KOODI HOTTA
IMECC-UNICAMP

RESUMO. Neste artigo propomos modelos de fatores latentes para realizar a modelagem conjunta de curvas de juros em múltiplos mercados, generalizando diversos modelos existentes na literatura de estimação da estrutura a termo de taxas de juros. Os modelos propostos não precisam assumir as restrições usuais de estimação e identificação, e assim possibilitam o uso de estruturas mais flexíveis com a incorporação de fatores latentes adicionais, volatilidade estocástica e a imposição de consistência com não-arbitragem.

A eliminação destas restrições é possível através da metodologia de estimação Bayesiana através de Markov Chain Monte Carlo (MCMC). Esta metodologia permite obter intervalos de credibilidade exatos para os parâmetros, fatores latentes e previsões, e também permite tratar os problemas de identificação e dimensionalidade existentes na estimação de modelos multimercados. Realizamos uma aplicação com a modelagem conjunta de curvas de Cupom Cambial e Eurodólares, realizando um procedimento extensivo de comparação de modelos e mostrando o potencial preditivo e prático dos modelos propostos.

1. INTRODUÇÃO

A modelagem da estrutura a termo de taxas de juros é um ponto fundamental na gestão de ativos financeiros. Uma enorme literatura tem se desenvolvido tentando obter formas mais precisas para a modelagem, previsão e precificação de instrumentos financeiros a partir da curva de juros. Dentro destas abordagens uma parte importante da literatura está baseada na idéia de que a evolução dinâmica da curva de juros pode ser descrita utilizando um conjunto de fatores dinâmicos que determinam a evolução dos prêmios de risco para as diversas maturidades observadas. A forma mais comum de tratar estes fatores é através de uma representação utilizando variáveis de estado latentes, ou seja, como variáveis não diretamente observadas¹.

Estes fatores latentes tem o objetivo de sumarizar todo o conjunto de variáveis relevantes que determinem os movimento nas curvas de juros. Metodologias para a extração destes fatores latentes podem partir de mecanismos puramente estatísticos, como a decomposição de componentes principais introduzida em [Litterman & Scheinkman, 1991], onde os fatores latentes são interpretados como componentes de nível, inclinação e curvatura.

Estes fatores latentes também podem ser identificados através de metodologias de precificação por equilíbrio, como os modelos de *short-rate* de [Vasicek, 1977] e [Cox *et al.*, 1985], pertencentes a classe de modelos afins (*Affine Diffusions*, e. g. [Dai & Singleton, 2000]). Estes modelos de equilíbrio também podem ser colocados em um framework geral baseado em condições de não-arbitragem através da formulação de Heath-Jarrow-Morton ([Heath *et al.*, 1992]), que determina a evolução das taxas forward como um processo estocástico de dimensão infinita.

Entretanto todas estas abordagens tem um sucesso parcial na modelagem empírica da evolução dinâmica da estrutura a termo de taxas de juros. Os modelos de equilíbrio e os modelos

Esta versão - Abril 2009. Endereço de contato - Ibmecc São Paulo - Rua Quatá 300, 04546-042, São Paulo, SP. Brasil. email - Márcio Laurini - marciopl@isp.edu.br - Luiz Koodi Hotta - hotta@ime.unicamp.br.

¹Para referências sobre a modelagem da estrutura a termo de taxas de juros veja por exemplo [Brigo & Mercurio, 2006] para aspectos relacionados a precificação de instrumentos financeiros e [Singleton, 2006] sobre a estimação de modelos de estrutura a termo de taxas de juros.

afins, embora possuam propriedades analíticas importantes como a existência de fórmulas fechadas para precificação de ativos, são caracterizadas por um ajuste bastante insatisfatório das taxas observadas, bem como das previsões derivadas destes modelos. Uma dificuldade adicional é que a estimação econométrica destes modelos em geral sofre de problemas de máximos locais e identificação, como aponta. [Duffe, 2002]. Os modelos de não-arbitragem são calibrados de forma a replicar com perfeição a curva de juros observada no mercado através de matching usando preços de títulos, mas esta calibração é de natureza cross-section e não permite previsões para curvas futuras e somente a precificação de instrumentos derivativos, e estes modelos são recalibrados diariamente utilizando instrumentos observados na curva de juros.

Tendo como objetivo principal a previsão da estrutura a termo das taxas de juros, [Diebold & Li, 2006] propõem um modelo dinâmico utilizando a forma paramétrica para a curva de juros proposta por [Nelson & Siegel, 1987]. O modelo proposto por [Diebold & Li, 2006] é uma extensão dinâmica da parametrização de [Nelson & Siegel, 1987], interpretando este modelo como um modelo de fatores latentes. Nesta generalização cada parâmetro do ajuste cross-section do modelo de Nelson-Siegel é tratado como um fator latente, e através da modelagem e previsão deste fator latente é possível obter previsões para toda a estrutura a termo de taxas de juros. Os resultados obtidos por [Diebold & Li, 2006] mostram que esta formulação obtém ajuste e poder preditivo superiores as demais metodologias de modelagem da curva de juros, tornando este modelo a referência padrão em previsão de estrutura a termo.

Além dos melhores resultados no ajuste e previsão das curvas de juros, o modelo proposto em [Diebold & Li, 2006] era atrativo pela facilidade de sua implementação. Com algumas restrições sobre o espaço paramétrico, este modelo poderia ser estimado utilizando apenas estimação por Mínimos Quadrados Ordinários, enquanto que os demais modelos utilizados necessitam de ferramentas de estimação mais complexas como o Filtro de Kalman (e.g. [Duffe, 2002]) ou métodos de estimação como o Método de Momentos Simulados, utilizado na estimação de modelos Afins em [Dai & Singleton, 2000]. Além de simplificar a sua implementação, as restrições impostas no modelo de [Diebold & Li, 2006] eram necessárias para evitar os problemas comuns existentes na estimação de modelos de estrutura a termo de taxas de juros, como os problemas já citados de máximos locais e não-identificação.

Com base no sucesso obtido na sua extensão dinâmica da curva de Nelson-Siegel, [Diebold *et al.*, 2008] propuseram uma generalização deste modelo para ajustar múltiplas curvas de juros simultaneamente, utilizando uma metodologia que consistia na construção de fatores latentes ligados a uma *curva de juros global* não diretamente observada. No modelo de [Diebold *et al.*, 2008] a curva de juros de cada mercado seria obtida como um deslocamento linear da curva global de juros mais um fator idiossincrático, através destes fatores latentes. É importante notar que a formulação de [Diebold *et al.*, 2008] é a primeira tentativa de um modelo que permita capturar a dinâmica de diversas estruturas a termo de forma simultânea. Esta formulação também foi adotada para modelar curvas de juros de países emergente em [Morita & Bueno, 2008], mostrando a aplicabilidade geral deste modelo.

No entanto o modelo proposto por [Diebold *et al.*, 2008] utiliza uma série de restrições em sua formulação. Devido ao elevado número de parâmetros envolvido na estimação do modelo global, [Diebold *et al.*, 2008] utilizam uma especificação bastante limitada para o formato geral da curva de juros em cada mercado. Ao invés de utilizar uma formulação completa do modelo de Nelson-Siegel com nível, inclinação e curvatura, eles utilizam apenas os componentes de nível e inclinação, tornando o ajuste para curvas de juros observadas bastante limitado, embora seja importante notar que o objetivo deste modelo não era primariamente de ajuste ou previsão, mas sim de verificar a existência de um fator global influenciando os movimentos da estrutura a termo nos mercados mais importantes.

Neste modelo também se utiliza uma restrição de que o parâmetro que define o padrão de decaimento da curva de juros seja mantido constante, o que prejudica bastante o ajuste do modelo. Outros problemas nesta formulação se referem aos procedimentos de estimação, utilizando um procedimento em dois estágios que não permite obter medidas como intervalos de confiança exatos para parâmetros do modelo e para as previsões da curva de juros. Outros problemas estão relacionados a identificação do modelo, isto é, obter condições para um único vetor de parâmetros

que defina o máximo da função de verossimilhança utilizada na estimação do modelo. Outros problemas nesta formulação estão relacionados ao pressuposto de uma volatilidade condicional constante, o que contraria um dos fatos estilizados na modelagem de curvas de juros.

Adicionalmente a formulação proposta em [Diebold *et al.*, 2008] também sofre de uma crítica fundamental existente sobre o modelo original de [Diebold & Li, 2006], que era a inconsistência do modelo com condições de não-arbitragem. Esta limitação original do modelo de [Diebold & Li, 2006] foi resolvida em [Christensen *et al.*, 2007, Christensen *et al.*, 2008], que mostram que embora a formulação original do modelo de [Diebold & Li, 2006] seja incompatível com condições de não-arbitragem, é possível trabalhar com uma forma aproximada deste modelo que é livre de arbitragem, reparametrizando o modelo de [Diebold & Li, 2006] como um modelo afim de estrutura a termo e obtendo um termo de correção que permite incorporar as condições de não-arbitragem neste contexto.

Uma dificuldade encontrada é que na formulação livre de arbitragem para a curva de Nelson-Siegel proposta por [Christensen *et al.*, 2007] é necessário que o modelo seja parametrizado com fatores de nível, inclinação e curvatura e na generalização colocada em [Christensen *et al.*, 2008] é necessário que o modelo tenha mais dois fatores de curvatura e inclinação. Desta forma a restrição imposta de somente ter fatores de nível e inclinação no modelo de fatores globais de [Diebold *et al.*, 2008] impede que sejam adotada estas parametrizações livres de arbitragem.

Neste artigo propomos modelos de fatores latentes para curvas de juros em múltiplos mercados, generalizando os modelos propostos por [Diebold & Li, 2006], [Diebold *et al.*, 2008] e [Christensen *et al.*, 2008]. Os modelos propostos não precisam assumir as restrições impostas originalmente nas formulações originais. Trabalhamos com uma forma paramétrica mais geral do que a adotada nestes artigos, utilizando a especificação proposta por [Svensson, 1994] com nível, inclinação e dois fatores de curvatura e também a forma funcional adotada em [Björk & Christensen, 1999] e [Christensen *et al.*, 2008] com fatores adicional de nível e curvatura. Estas formas paramétricas mais flexíveis permitem uma ajuste muito mais preciso para a estrutura a termo de taxas de juros. Outra extensão proposta é um fator adicional de volatilidade estocástica, permitindo capturar o fato estilizado da presença de volatilidade condicional observado em taxas de juros (e.g. [Chan *et al.*, 1992]).

A metodologia de estimação proposta aqui também permite que possamos tratar os parâmetros que definem o padrão de decaimento da curva de juros em cada instante do tempo como um fatores latentes adicionais e variantes no tempo, permitindo um melhor ajuste da curva de juros. A eliminação destas restrições é possível através de metodologia de estimação Bayesiana e um procedimento de estimação através de mecanismos de Markov Chain Monte Carlo (MCMC). Esta metodologia de estimação adicionalmente também permite obter intervalos de credibilidade exatos para os parâmetros e previsões do modelo, e representa uma forma de contornar os problemas de estimação existentes nas metodologias tradicionais de estimação de modelos de estrutura a termo de taxas de juros como mínimos quadrados não-lineares (e.g [Svensson, 1994]) ou máxima verossimilhança utilizando o filtro de Kalman [Duffe, 2002], em especial os problemas de máximos locais e não identificação existentes nestas formulações.

Esta metodologia também permite tratar o problema da dimensionalidade existente na estimação de modelos multimercados através de um mecanismo conhecido como *Bayesian Shrinkage*, que permite eliminar automaticamente os parâmetros redundantes do modelo. Finalmente implementamos para o modelo de múltiplos mercados as condições de não-arbitragem formuladas por [Christensen *et al.*, 2008], generalizando estas condições para o caso de múltiplos mercados. Desta forma as generalizações propostas tratam de todos os problemas apontados nas formulações originais dos modelos de [Diebold & Li, 2006] e [Diebold *et al.*, 2008].

Este artigo está estruturado da seguinte forma - nas seções 2 e 3 revisamos os modelos originais de [Diebold & Li, 2006] e [Diebold *et al.*, 2008] e discutimos os problemas existentes nestas formulações. Na seção 4 mostramos as as extensões propostas contornar estes problemas. A seção 5 discute a implementação das condições de não-arbitragem, e a seção 6 mostra o procedimento de estimação Bayesiano por MCMC utilizado. Nas seções 8 e 7 discutimos como a estimação Bayesiana permite tratar os problemas de identificação e da dimensionalidade do vetor de parâmetros. Na seção 9 mostramos uma aplicação empírica dos modelos propostos, ajustando modelos conjuntos para as curvas de Cupom Cambial no Brasil e a curva de Eurodólares. Nesta seção realizamos

uma comparação extensiva de todos os modelos propostos no artigo e também implementamos um procedimento novo na literatura que permite verificar a validade da imposição das condições de não-arbitragem nestas modelos de estrutura a termo de taxas de juros. As considerações finais estão na seção 10.

2. MODELO DE DIEBOLD & LI

Entre os modelos utilizados para modelagem da estrutura a termo das taxas de juros, o modelo proposto por [Diebold & Li, 2006] é bastante difundido no mercado, pela sua simplicidade de implementação e por sua performance preditiva superior. Este modelo é baseado na formulação proposta por [Nelson & Siegel, 1987] para o ajuste cross-section (dia a dia) da curva de juros. A curva de [Nelson & Siegel, 1987] é representada como:

$$(2.1) \quad y_t(m) = \beta_1 + \beta_2 \frac{1 - e^{-m/\tau}}{m/\tau} + \beta_3 \left[\frac{1 - e^{-m/\tau}}{m/\tau} - e^{-m/\tau} \right] + \epsilon_m$$

onde $y_t(m)$ são as taxas observadas em uma certa data t para o vetor de maturidades m , e $\beta_1, \beta_2, \beta_3$ e τ são parâmetros. Os parâmetros são interpretáveis - β_1 representa o componente de longo-prazo, β_2 componente de curto prazo, β_3 um componente de médio prazo e τ é um parâmetro que controla o padrão de decaimento da curva de juros. Os parâmetros $\beta_1, \beta_2, \beta_3$ também podem ser interpretados como decomposições de nível, inclinação e curvatura de acordo com a terminologia de desenvolvida por [Litterman & Scheinkman, 1991]. Este modelo é uma forma parcimoniosa de ajustar a curva de juros, e é capaz de reproduzir diversos fatos estilizados sobre o formato da curva de juros no tempo.

[Diebold & Li, 2006] tornam o modelo de Nelson-Siegel dinâmico, interpretando os parâmetros como fatores dinâmicos. Este modelo pode ser formulado através de uma equação de observação para a curva de juros dada por:

$$(2.2) \quad y_t(m) = \beta_{1t} + \beta_{2t} \frac{1 - e^{-m/\tau}}{m/\tau} + \beta_{3t} \left[\frac{1 - e^{-m/\tau}}{m/\tau} - e^{-m/\tau} \right] + \epsilon_{ym}$$

e um sistema determinando a evolução dos fatores latentes como um vetor autoregressivo vetorial de primeira ordem:

$$(2.3) \quad \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \Phi \begin{bmatrix} \beta_{1t-1} \\ \beta_{2t-1} \\ \beta_{3t-1} \end{bmatrix} + \epsilon_{\beta t}$$

onde Φ é a matriz de parâmetros deste vetor autoregressivo. A estimação do modelo é normalmente realizada através de um procedimento em dois estágios. O primeiro estágio é a estimação da equação 2.2 para cada dia observado. Esta estimação é realizada por Mínimos Quadrados Ordinários, assumindo que o parâmetro de decaimento τ é fixo e conhecido, estimando os fatores latentes $\beta_{1t}, \beta_{2t}, \beta_{3t}$ para cada período de tempo t . O segundo estágio consiste na estimação por Mínimos Quadrados Ordinários da matriz de parâmetros Φ do vetor autoregressivo usando os parâmetros β_{1t}, β_{2t} and β_{3t} , estimados no primeiro estágio. Previsões para o modelo são obtidas acoplando as previsões para os fatores latentes t dias a frente na equação de Nelson-Siegel (2.2).

Como é possível notar a estimação e previsão no modelo de [Diebold & Li, 2006] é extremamente simples, sendo possível de ser implementada em qualquer software padrão de econometria. No entanto esta formulação simplificada pode ser criticada em vários aspectos. Pode ser muito restritivo considerar o parâmetro τ constante para curvas instáveis, como curvas de países emergentes. Este parâmetro captura o decaimento médio da curva de juros, e com alterações no formato da curva este parâmetro deve se modificar. Outro ponto importante é que a especificação paramétrica adotada derivada da forma funcional do modelo de Nelson-Siegel não permite capturar curvas com formatos mais complicados, como curvas que possuem mais de uma mudança na inclinação e na curvatura.

Outros pontos importantes se referem as propriedades dos estimadores neste procedimento de estimação em dois estágios. O primeiro ponto é que a estimação só é consistente na escolha de um parâmetro τ correto. Também é importante notar que a distribuição dos estimadores neste contexto não é usual, já que a estimação no segundo estágio é baseada em uma série construída através de um primeiro estágio. Isso também afeta a construção dos intervalos de confiança para as previsões da curva de juros derivada deste modelo. Note também que existe uma perda de eficiência na estimação em dois estágios, já que a estimação dos fatores latentes é realizada dia a dia, desconectada da estrutura de vetor autoregressivo assumida na equação 2.3. Uma forma alternativa de realizar esta estimação seria usando máxima verossimilhança através do Filtro de Kalman, já que o sistema formado pelas equações 2.2e 2.3 já está em uma formulação em espaço de estados, mas esta estimação continua sofrendo de problemas de máximos locais e não-identificação, como é comum na estimação de modelos de estrutura a termo usando o Filtro de Kalman (e.g. [Duffe, 2002]). Outro problema fundamental é que a formulação original do modelo de [Diebold & Li, 2006] não é consistente com o princípio de não-arbitragem. A curva de Nelson-Siegel utilizada não admite uma representação livre de arbitragem, pode ser visto por exemplo em [Björk & Christensen, 1999], [Filipovic, 1999, Filipovic, 2001] e [Diebold *et al.*, 2005].

3. EXTENSÕES PARA MÚLTIPLOS MERCADOS

O modelo de Diebold-Li é um modelo dinâmico para a curva de apenas um mercado, mas é possível generalizar esta formulação para modelar diversas curvas de juros simultaneamente. Esta generalização foi proposta em [Diebold *et al.*, 2008]. Denotando a curva observada para o mercado i em função de um vetor de maturidades² τ por $y_{it}(\tau)$, temos que neste modelo a dinâmica dos yields é dada por uma versão restrita da curva de Nelson-Siegel, com apenas os fatores de Nível e Inclinação:

$$(3.1) \quad y_{it}(\tau) = l_{it} + s_{it} \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + v_{it}(\tau)$$

onde l_{it} na notação de [Diebold *et al.*, 2008] representa o componente de nível no período t para o país i , s_{it} o componente de inclinação para este mesmo país em cada período t , e v_{it} é um componente de choques para a equação de observação das taxas. Para especificar a dinâmica completa do modelo é necessário especificar a evolução dos fatores latentes de nível e inclinação para cada país. Na especificação proposta em [Diebold *et al.*, 2008] a idéia é que existam os chamados *fatores globais* determinados por uma curva não-observada y_{gt} na forma:

$$(3.2) \quad y_{gt}(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + V_{it}(\tau)$$

e a dinâmica dos *fatores latentes globais* L_t e S_t é dada pela seguinte dinâmica autoregressiva:

$$(3.3) \quad \begin{pmatrix} L_t \\ S_t \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} L_{t-1} \\ S_{t-1} \end{pmatrix} + \begin{pmatrix} U_t^l \\ U_t^s \end{pmatrix}$$

Para determinar os componentes de inclinação e curvatura é assumido que a curva de cada país é uma modificação linear da curva global mais um componente idiossincrático. As curvas locais são dadas por:

$$(3.4) \quad \begin{aligned} l_{it} &= \alpha_i^l + \beta_i^l L_t + \varepsilon_{it}^l \\ s_{it} &= \alpha_i^s + \beta_i^s L_t + \varepsilon_{it}^s \end{aligned}$$

²Nesta exposição deste modelo seguimos a notação original do artigo de [Diebold *et al.*, 2008], que denota o vetor de maturidades por τ e o parâmetro de decaimento por λ , enquanto que a notação utilizada nos demais modelos apresentados usa m para denotar o vetor de maturidades e τ para os parâmetros de decaimento, e utilizamos a especificação original de [Nelson & Siegel, 1987] para o parâmetro τ , enquanto que [Diebold & Li, 2006], [Diebold *et al.*, 2008] utilizam o fator $\lambda = 1/\tau$.

e os componentes idiossincráticos são dados por processos autoregressivos de primeira ordem:

$$(3.5) \quad \begin{pmatrix} \varepsilon_{it}^l \\ \varepsilon_{it}^s \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{it-1}^l \\ \varepsilon_{it-1}^s \end{pmatrix} + \begin{pmatrix} u_t^l \\ u_t^s \end{pmatrix}$$

A estimação deste modelo poderia ser realizada em princípio utilizando máxima verossimilhança através da decomposição do erro de previsão usando o filtro de Kalman, notando que neste caso temos as variáveis latentes adicionais que representam os fatores globais. Mas devido a dimensão do problema para o caso multimercado e os problemas usuais de estimação, como os problemas de identificação e a possibilidade de máximos locais, a estimação do modelo de [Diebold *et al.*, 2008] é realizada em dois estágios.

No primeiro estágio a curva é obtida para cada país por Mínimos Quadrados Ordinários, assumindo novamente que o parâmetro que controla o decaimento da curva é mantido constante e não estimado. Com os fatores obtidos para cada país é realizado um segundo estágio utilizando Markov Chain Monte Carlo para obter os demais parâmetros e fatores latentes. A estimação também é realizada impondo algumas restrições, como assumir que a matriz de parâmetros nos processos autoregressivos dos fatores locais sejam diagonais. Embora este procedimento tenha uma finalidade operacional, é difícil obter uma interpretação estatística dos resultados, já que a estimação em parte do modelo é frequentista e em parte Bayesiana. E novamente temos o mesmo problema de como construir intervalos de confiança para parâmetros e previsões com este procedimento em dois estágios, e os procedimentos de estimação por MCMC utilizam somente distribuições conjugadas e especificações lineares, não utilizando a informação total existente nas curvas de juros.

Este procedimento apresenta limitações semelhantes as da estimação original do modelo de Diebold-Li, mas agravadas pela dimensionalidade e heterogeneidade do modelo e dos formatos das curvas de juros entre distintos mercados. Um primeiro ponto importante é notar que a fixação do parâmetro de decaimento da curva pode limitar bastante o ajuste do modelo. Mercados distintos podem ter fatores de decaimento bastante diferentes, e como já colocado assumir que estes parâmetros são fixos no tempo pode ser extremamente limitante.

Outro ponto importante é que a restrição de assumir apenas a existência de fatores de nível e inclinação também limita o ajuste possível do modelo. Existe uma grande literatura citando os ganhos de ajuste dados pela inclusão de fatores adicionais de curvatura, como o modelo original de [Svensson, 1994] e os modelos de [Björk & Christensen, 1999] que adicional mais fatores de inclinação e de curvatura, aumentando de forma sensível o ajuste. Também é fundamental notar que nesta especificação não é possível utilizar as especificações livres de arbitragem propostas em [Christensen *et al.*, 2007, Christensen *et al.*, 2008], já que nestas formulações cada componente de inclinação tem que estar casado com componente de curvatura com a mesma taxa de reversão a média, e assim esta formulação sem componentes de curvatura não pode ser tornada livre de arbitragem.

Outra crítica fundamental é que na especificação dinâmica adotada na equação 3.4 temos que a curva de cada país é um deslocamento da curva global mais um fator idiossincrático. Note que nesta formulação não existe uma interdependência direta entre as curvas de juros, e assim o modelo não permite identificar diretamente as possíveis interações entre os fatores latentes de distintos mercados. Uma interpretação esperada seria verificar se por exemplo deslocamentos no nível de um mercado afetam o nível dos demais mercados. Note que nesta formulação isto é realizado apenas de forma indireta por modificações no fator global, mas esse efeito direto não é possível de ser observado.

4. MODELOS PROPOSTOS

Para tratar dos problemas existentes nas formulações originais dos modelos de [Diebold & Li, 2006, Diebold *et al.*, 2008] nós utilizamos generalizamos o framework Bayesiano de fatores latentes proposto em [Laurini & Hotta, 2008], mas generalizado para o caso de mais de uma curva de juros e também com a adição da correção de não-arbitragem proposta em [Christensen *et al.*, 2008]. Os modelos propostos podem ser classificados em 3 classes - a primeira classe de modelos é uma generalização da estrutura de fatores latentes, aumentando o vetor de estado de forma

a incluir interações com os demais fatores latentes, em especial os fatores latentes dos demais países. A segunda classe de modelos é uma generalização da estrutura de fatores globais de [Diebold *et al.*, 2008], com a inclusão de componentes de curvatura, dupla curvatura e inclinações adicionais. A terceira classe contém as modificações necessárias para tornar as duas classes anteriores livres de arbitragem, utilizando a aproximação de modelo afim proposta em [Christensen *et al.*, 2007, Christensen *et al.*, 2008].

A estrutura comum entre as duas primeiras classes é dada pela formulação mais flexível da equação de observação. Adotamos como estrutura básica a generalização dinâmica da forma paramétrica proposta por [Svensson, 1994], que consiste em uma equação com um fator de nível, inclinação e dois fatores de curvatura na forma:

$$(4.1) \quad y_t(m) = \beta_{1t} + \beta_{2t} \frac{1 - e^{-m/\tau_{1t}}}{m/\tau_{1t}} + \beta_3 \left[\frac{1 - e^{-m/\tau_{1t}}}{m/\tau_{1t}} - e^{-m/\tau_{1t}} \right] \\ + \beta_{4t} \left[\frac{1 - e^{-m/\tau_{2t}}}{m/\tau_{2t}} - e^{-m/\tau_{2t}} \right] + \sigma_t \eta_t$$

aonde assumimos:

$$(4.2) \quad \eta_t \sim IID(0, 1) \text{ e } \eta_t \perp \eta_s \quad \forall t \neq s$$

Nos modelos livre de arbitragens utilizamos a especificação com um fator adicional de inclinação e outro de curvatura sobre a representação dada pela equação 4.1, detalhada na seção 5. Sobre este modelo básico também adotamos a generalização proposta em [Laurini & Hotta, 2008] de tornar os fatores de decaimento τ_1 e τ_2 variantes no tempo, tratando estes parâmetros como fatores latentes adicionais, utilizando uma estrutura autoregressiva de primeira ordem para todos os fatores latentes dada por:

$$(4.3) \quad \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \\ \beta_{4t} \\ \tau_{1t} \\ \tau_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{\beta_1} \\ \mu_{\beta_2} \\ \mu_{\beta_3} \\ \mu_{\beta_4} \\ \mu_{\tau_1} \\ \mu_{\tau_2} \end{bmatrix} + \Phi \begin{bmatrix} \beta_{1t-1} \\ \beta_{2t-1} \\ \beta_{3t-1} \\ \beta_{4t-1} \\ \tau_{1t-1} \\ \tau_{2t-1} \end{bmatrix} + \epsilon_t$$

onde a princípio a matriz Φ é uma matriz completa, e assim cada fator latente no período t depende dos demais fatores latentes no período $t-1$ mais um intercepto μ . Outra generalização existente neste modelo é a possibilidade de um fator de Volatilidade Estocástica σ_t , cuja dinâmica é dada por:

$$(4.4) \quad \ln \sigma_t^2 = \phi_0 + \phi_1 \ln \sigma_{t-1}^2 + v_t$$

Este fator permite capturar uma estrutura de volatilidade condicional nas taxas de juros, permitindo capturar este fato estilizado (e.g. [Chan *et al.*, 1992], [Lund & Andersen, 1997]). Este componente de volatilidade estocástica tem uma função adicional, que é a de possivelmente evitar uma variação excessiva nos fatores latentes do modelo. Um resultado conhecido na literatura Bayesiana é que é possível escrever um modelo de regressão com coeficientes aleatórios em um modelo de regressão com coeficientes fixos através da inclusão de um componente de heterocedasticidade condicional (e.g. [Bauwens *et al.*, 1999]). Sobre este ponto também é interessante observar a crítica realizada por [Sims, 2001] a um modelo de parâmetros variantes no tempo proposto por [Cogley & Sargent, 2001] para observar variações na política monetária. Sims aponta que a variação nos parâmetros observada no modelo de [Cogley & Sargent, 2001] poderia ser gerada por uma estrutura de volatilidade condicional não controlada no modelo. Desta forma este componente de volatilidade condicional tenta evitar este problema de variação excessiva nos fatores latentes do modelo.

Um ponto importante é que na classe de modelos com correções para não-arbitragem é necessário manter as estruturas de volatilidade condicional e de parâmetros de decaimento constantes no tempo, e assim estas duas extensões não podem ser utilizadas. Estas extensões tornariam o modelo não pertencente a classe de modelos afins, e assim a aproximação proposta em [Christensen *et al.*, 2007, Christensen *et al.*, 2008] não se aplicaria.

Nas seções seguintes definimos as características particulares das 3 classes de modelos propostas no artigo.

4.1. Modelo de Fatores Latentes Generalizados. A primeira classe de modelos utiliza uma generalização do modelo de Diebold-Li, expandindo os vetores latentes para incluir interações entre os fatores latentes definindo as curvas dos diversos mercados. . Nesta classe definimos a curva observada $y_t^i(m)$ para o país i através da representação de Svensson:

$$(4.5) \quad y_t^i(m) = \beta_{1t}^i + \beta_{2t}^i \frac{1 - e^{-m/\tau_{1t}^i}}{m/\tau_{1t}^i} + \beta_3^i \left[\frac{1 - e^{-m/\tau_{1t}^i}}{m/\tau_{1t}^i} - e^{-m/\tau_{1t}^i} \right] \\ + \beta_{4t}^i \left[\frac{1 - e^{-m/\tau_{2t}^i}}{m/\tau_{2t}^i} - e^{-m/\tau_{2t}^i} \right] + \sigma_t^i \eta_t^i$$

A generalização do vetor de fatores latentes para o caso multimercado é dada pela seguinte representação:

$$(4.6) \quad \beta_{kt}^i = \Phi_i \beta_{kt-1}^i + \Phi_j \beta_{kt-1}^j + \epsilon_{kt}$$

$$(4.7) \quad \tau_{kt}^i = \theta_i \tau_{kt-1}^i + \theta_j \tau_{kt-1}^j + \nu_{kt}$$

$$(4.8) \quad \ln \sigma_t^i = \gamma_i \ln \sigma_{t-1}^i + \gamma_j \ln \sigma_{t-1}^j + \xi_{kt}$$

onde β_{kt}^i representa o fatores latentes de nível, inclinação, curvatura e dupla curvatura para o mercado i , com uma estrutura análoga a equação 4.3, mas com a inclusão de β_{kt-1}^j , que representa os fatores para o mercado j , que possuem uma representação equivalente. Da mesma forma temos os fatores de decaimento τ_{kt}^i para os diversos países e a estrutura de volatilidade estocástica σ_t^i para cada país i . Note que nesta representação cada fator latente de um mercado é influenciado pelos fatores dos demais países, permitindo introduzir uma interação entre as diversas curvas de juros, como discutido na seção 3. Para completar o modelo adotamos a seguinte estrutura de covariância para os parâmetros de cada mercado:

$$\sum_{\eta, \epsilon, \nu}^i = \begin{bmatrix} \sigma_\eta^{i2} & 0 & 0 \\ 0 & \Omega_\epsilon^i & 0 \\ 0 & 0 & \sigma_\nu^i \end{bmatrix}$$

A matriz $\sum_{\eta, \epsilon, \nu}^i$ é a matriz expandida de variância-covariância dos parâmetros do modelo para cada país; σ_η^{i2} é a variância da equação de medida ; Ω_ϵ^i é a matriz de covariância entre os fatores latentes; e σ_ν^i é a variância do processo de volatilidade estocástica. Esta matriz é bloco-diagonal, exceto pela sub-matriz de componentes Ω_ϵ^i , que podem ser correlacionados.

4.2. Modelo Global Generalizado. A segunda classe de modelos é uma generalização do modelo de fatores globais de [Diebold *et al.*, 2008]. Neste caso não adotamos as restrições deste artigo, e utilizamos uma representação completa para a estrutura paramétrica da curva de juros observada em cada país, usando uma representação análoga a curva de Svensson. Seguindo a notação de [Diebold *et al.*, 2008] nós temos que a curva de cada país é dada por:

$$(4.9) \quad y_t^i(m) = l_{it} + s_{it} \frac{1 - e^{-m/\tau_{1t}^i}}{m/\tau_{1t}^i} + c_{1it} \left[\frac{1 - e^{-m/\tau_{1t}^i}}{m/\tau_{1t}^i} - e^{-m/\tau_{1t}^i} \right] \\ + c_{2it}^i \left[\frac{1 - e^{-m/\tau_{2t}^i}}{m/\tau_{2t}^i} - e^{-m/\tau_{2t}^i} \right] + \sigma_t^i \eta_t^i$$

onde l_{it} é o nível do país i , s_{it} é a inclinação e c_{1it} e c_{2it} são os dois fatores de curvatura, sendo que todos estes fatores evoluem em t . Nesta representação temos que τ_{1t}^i e τ_{2t}^i são os fatores de decaimento para cada país i , e também são variantes no tempo.

Para completar a especificação do modelo, generalizamos a estrutura de fatores globais utilizada por [Diebold *et al.*, 2008]. Nesta estrutura cada fator latente de nível, inclinação e curvaturas são uma função linear do fator global equivalente. Esta representação é escrita como:

$$(4.10) \quad l_{it} = \alpha_i^l + \beta_i^l L_t + \varepsilon_{it}^l$$

$$(4.11) \quad s_{it} = \alpha_i^s + \beta_i^s S_t + \varepsilon_{it}^s$$

$$(4.12) \quad c_{1it} = \alpha_i^{c1} + \beta_i^{c1} C_{1t} + \varepsilon_{it}^{c1}$$

$$(4.13) \quad c_{2it} = \alpha_i^{c2} + \beta_i^{c2} C_{2t} + \varepsilon_{it}^{c2}$$

$$(4.14) \quad \tau_{1it} = \alpha_i^{\tau1} + \beta_i^{\tau1} \tau_{g1t} + \varepsilon_{it}^{\tau1}$$

$$(4.15) \quad \tau_{2it} = \alpha_i^{\tau2} + \beta_i^{\tau2} \tau_{g2t} + \varepsilon_{it}^{\tau2}$$

e os α e β representam parâmetros (loadings) a serem estimados e o vetor de fatores latentes globais $(L_t, S_t, C_{1t}, C_{2t}, \tau_{g1t}, \tau_{g2t})$ evolui como um vetor autoregressivo de primeira ordem, generalizando a estrutura da equação 3.3. Também assumimos que os componentes idiossincráticos para os fatores latentes de cada mercado seguem uma estrutura autoregressiva de primeira ordem, de acordo com a especificação geral dada pela equação 3.5, mas para este vetor generalizado de fatores latentes.

5. NÃO-ARBITRAGEM

Nas especificações discutidas até agora temos representações basicamente estatísticas, i.e., embora os fatores latentes tenha uma interpretação como componentes de nível, inclinação e curvatura, mesmo em modelos afins esta interpretação é uma aproximação, como mostra [Almeida, 2005]. Estas representações são apenas ferramentas de ajuste e previsão da curva de juros, sem uma justificativa teórica ou estrutural completa. Neste ponto a principal deficiência destes modelos é a compatibilidade com o princípio de precificação por Não-Arbitragem. O resultado fundamental da precificação por Não-Arbitragem, conhecido como o Teorema Fundamental da Precificação de Ativos, estabelece que um mercado é livre de arbitragem se e somente se existe (pelo menos uma) medida de Probabilidade Q equivalente a medida física P tal que a seqüência descontada de retornos do ativo ajustados para risco sejam um semi-martingale nesta medida Q (e.g. [Harrison & Kreps, 1979, Harrison & Pliska, 1981, Delbaen & Schachermayer, 1994]).

Consistência com o princípio de não-arbitragem é uma propriedade fundamental em finanças, já que ela estabelece que o retorno de ativo deve ser consistente com seu nível de risco, e assim não devem existir de forma sistemática lucros certos sem risco. Em mercados de ativos grandes e com

elevada liquidez o princípio de não-arbitragem deve ser alcançado pela atuação de traders racionais. Na modelagem da estrutura a termo de taxas de juros o princípio geral de não-arbitragem pode ser observado dentro do framework geral proposto por [Heath *et al.*, 1992]. Uma curva é consistente com não-arbitragem se ela pode ser projetada no espaço de todas as curvas livres de arbitragem na medida martingale equivalente, e de uma forma geral ela deve estar contida dentro de uma variedade estocástica gerada pela estrutura de Heath-Jarrow-Morton, como mostrado em [Filipovic, 2001].

O problema é que as curvas geradas pelos modelos de Nelson-Siegel nunca são consistentes com não-arbitragem, e existe apenas uma restrição sobre o modelo de Svensson consistente com não-arbitragem, mas esta estrutura é limitada demais para utilização prática, como provado por [Filipovic, 1999]. Desta forma embora os modelos da classe Nelson-Siegel e Svensson e suas extensões dinâmicas tenham um bom ajuste empírico aos dados observados da estrutura a termo de taxas de juros, não seriam válidos em termos de uma precificação consistente com não arbitragem. No entanto a situação oposta também ocorre - a maioria dos modelos de não-arbitragem utilizados tem um ajuste ruim aos dados observados, como mostra [Duffe, 2002], existindo um aparente trade-off entre consistência com não-arbitragem e poder de ajuste e previsão. Mas evidências recentes mostram que com modificações adequadas na estrutura de modelos livres de arbitragem é possível obter poder preditivo adequado nestes modelos, como por exemplo em [Almeida & Vicente, 2008].

Embora não exista uma forma livre de arbitragem na classe de Nelson-Siegel-Svensson, algumas modificações permitem uma classe similar de modelos com a propriedade de não-arbitragem, como mostrado em [Christensen *et al.*, 2007] para a família de Nelson-Siegel e [Christensen *et al.*, 2008] para a família de Svensson. Para realizar esta correção para não-arbitragem [Christensen *et al.*, 2007, Christensen *et al.*, 2008] utilizam uma estrutura de modelos afins de estrutura a termo (*Affine Term Structure Models* - ATSM). A estrutura de modelos afins (ATSM) é bastante conveniente, já que apresenta propriedades analíticas interessantes, como por exemplo a existência de fórmulas fechadas para precificação de ativos e são caracterizados por uma estrutura comum que permite encompassar diversos modelos estudados na literatura, como mostram [Dai & Singleton, 2000].

Para caracterizar a estrutura dos modelos afins de estrutura a termo (ATSM) partimos da definição de um preço de um título zero cupom no período t com maturidade T na medida martingale equivalente Q , que deve ser dada por:

$$(5.1) \quad P(t, T) = E_t^Q \left[e^{-\int_t^T r_s ds} \right]$$

onde $r(t)$ representa a taxa instantânea de juros (*short rate*). Nesta classe de modelos $r(t)$ é uma função afim de um vetor de variáveis de estado (fatores latentes) não observados $Y(t)$:

$$(5.2) \quad r(t) = \delta_0 + \sum_{i=1}^N \delta'_i Y_i(t)$$

onde os δ s representam parâmetros e $Y_i(t)$ é uma chamada difusão afim (*Affine diffusion*) com a seguinte estrutura:

$$(5.3) \quad dY_i(t) = \kappa(\theta - Y_i(t))dt + \sum \sqrt{S(t)}dW(t)$$

com κ e θ parâmetros, $dW(t)$ é um movimento Browniano Padrão e $S(t)$ é uma matriz diagonal com i -ésimo elemento dado por:

$$(5.4) \quad S(t)_{ii} = \alpha_i + \beta'_i Y(t)$$

[Duffie & Kan, 1996] mostram que desta forma temos que o preço do título pode ser escrito como:

$$(5.5) \quad P(t, \tau) = e^{A(\tau) - B(\tau)' Y(t)}$$

onde $A(\tau)$ e $B(\tau)$ são dados pela a solução do seguinte sistema de equações diferenciais ordinárias::

$$(5.6) \quad \begin{aligned} \frac{dA(\tau)}{dt} &= -\theta' \kappa' B(\tau) + \frac{1}{2} \sum_{i=1}^N \left[\sum' B(\tau) \right]_i^2 \alpha_i - \delta_0 \\ \frac{dB(\tau)}{dt} &= -\kappa' B(\tau) + \frac{1}{2} \sum_{i=1}^N \left[\sum' B(\tau) \right]_i^2 \beta_i - \delta_y \end{aligned}$$

A grande vantagem da classe de modelos afins de estrutura a termo é que ela é bastante flexível, permitindo generalizar uma ampla gama de modelos de estrutura a termo utilizados na literatura, em especial na definição dos fatores latentes que podem ser bastante gerais, como mostram [Dai & Singleton, 2000] e [Diebold *et al.*, 2005].

Para obter uma representação livre de arbitragem para a família de modelos de estrutura a termo definidos pela curva de Svensson³, [Christensen *et al.*, 2008] utilizam uma estrutura de modelo afim, assumindo que a short rate é dada pela soma de 5 fatores latentes:

$$(5.7) \quad r(t) = X_t^1 + X_t^2 + X_t^3 + X_t^4 + X_t^5$$

e estes fatores latentes $X_t^1, X_t^2, X_t^3, X_t^4, X_t^5$ evoluem através do seguinte sistema de equações diferenciais estocásticas:

$$(5.8) \quad \begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \\ dX_t^4 \\ dX_t^5 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & -\lambda_1 & 0 \\ 0 & 0 & \lambda_1 & 0 & -\lambda_2 \\ 0 & 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \begin{bmatrix} \begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \\ \theta_4^Q \\ \theta_5^Q \end{pmatrix} \\ - \begin{pmatrix} X_t^1 \\ X_t^2 \\ X_t^3 \\ X_t^4 \\ X_t^5 \end{pmatrix} \end{bmatrix}$$

Neste modelo, em acordo com a equação 5.1, preços de títulos zero-cupom são dados pela seguinte expressão:

$$(5.9) \quad P(t, T) = E_t^Q [e^{\int_t^T r_u du}] = \exp(B^1(t, T)X + B^2(t, T)X_t^2 + B^3(t, T)X_t^3 + B^4(t, T)X_t^4 + B^5(t, T)X_t^5 + C(t, T))$$

onde os termos $B^i(t, T)$ e $C(t, T)$ são as únicas soluções para os seguintes sistemas de equações diferenciais ordinárias:

$$(5.10) \quad \begin{pmatrix} \frac{dB^1(t, T)}{dt} \\ \frac{dB^2(t, T)}{dt} \\ \frac{dB^3(t, T)}{dt} \\ \frac{dB^4(t, T)}{dt} \\ \frac{dB^5(t, T)}{dt} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & 0 \\ 0 & -\lambda_1 & 0 & \lambda_1 & 0 \\ 0 & 0 & -\lambda_2 & 0 & \lambda_2 \end{pmatrix} \begin{bmatrix} \begin{pmatrix} B^1(t, T) \\ B^2(t, T) \\ B^3(t, T) \\ B^4(t, T) \\ B^5(t, T) \end{pmatrix} \end{bmatrix}$$

$$(5.11) \quad \frac{dC(t, T)}{dt} = -B(t, T)' \kappa^Q \theta^Q - \frac{1}{2} \sum_{j=1}^5 \left(\sum' B(t, T) B(t, T)' \sum \right)_{j,j}$$

e os zero-cupom yields são dados por:

$$(5.12) \quad y(t, T) = X_t^1 + \frac{1 - e^{-\lambda_1(t-T)}}{\lambda_1(t-T)} X_t^2 + \frac{1 - e^{-\lambda_2(t-T)}}{\lambda_2(t-T)} X_t^3 +$$

$$\left[\frac{1 - e^{-\lambda_1(t-T)}}{\lambda_1(t-T)} - e^{-\lambda_1(t-T)} \right] X_t^4 + \left[\frac{1 - e^{-\lambda_2(t-T)}}{\lambda_2(t-T)} - e^{-\lambda_2(t-T)} \right] X_t^5 - \frac{C(t, T)}{T-t}$$

³A derivação para a família Nelson-Siegel é um caso particular, utilizando apenas os fatores latentes X_t^1, X_t^2 e X_t^3 , como pode ser visto em [Christensen *et al.*, 2007].

que pode ser interpretado como a reparametrização da curva de Svensson com a adição de um fator de correção para não-arbitragem dado pelo termo $-\frac{C(t,T)}{T-t}$, que é dado pela seguinte expressão⁴:

$$(5.13) \quad -\frac{C(t,T)}{T-t} = -\frac{1}{2} \frac{1}{T-t} \sum_{j=1}^5 \left(\sum B(s,T) B(s,T) \cdot \sum \cdot \right),$$

onde \sum é uma matriz com covariâncias dos fatores latentes. Este fator de correção é uma função é uma função das variâncias dos fatores latentes e também dos parâmetros de decaimento do modelo, que nesta formulação são assumidos constantes.

Mas a especificação do modelo de [Christensen *et al.*, 2008] é uma representação muito útil, já que qualquer forma afim de fatores latentes pode ser utilizada, permitindo por exemplo acrescentar variáveis macroeconômicas no vetor de fatores latentes ou então aumentar a estrutura de dependência nestes fatores, mantendo a consistência com não-arbitragem, que é a forma utilizada no presente artigo, incluindo a interação com os fatores latentes dos demais mercados de forma a generalizar a estrutura de fatores utilizada.

Um ponto importante é que para o uso da correção para não-arbitragem precisamos de uma estrutura com cinco fatores latentes, o que implica um fatores adicionais de inclinação e curvatura no modelo de Svensson, e assim a representação original do modelo de [Diebold *et al.*, 2008] com apenas fatores de nível e inclinação não pode ser tornada livre de arbitragem pela metodologia de [Christensen *et al.*, 2007, Christensen *et al.*, 2008] .

Para obter representações livre de arbitragem para os modelos de fatores latentes generalizados proposto na seção 4, aumentamos a dinâmica de fatores latentes pela inclusão dos fatores cruzados, i.e., cada fator latente em cada mercado depende dos fatores latentes do próprio mercado mais os fatores latentes dos demais mercados, na forma:

$$(5.14) \quad \beta_{kt}^i = \Phi_i \beta_{kt-1}^i + \Phi_j \beta_{kt-1}^j + \epsilon_{kt}$$

onde agora $k=1,2,3,4,5$ representando os cinco fatores necessários a correção por não-arbitragem, e agora a equação que descreve os yields de cada mercado é dada por:

$$(5.15) \quad y^i(t, T) = \beta_{1t}^i + \frac{1 - e^{-\lambda_{i1}(t-T)}}{\lambda_{i1}(t-T)} \beta_{2t}^i + \frac{1 - e^{-\lambda_{i2}(t-T)}}{\lambda_{i2}(t-T)} \beta_{3t}^i + \left[\frac{1 - e^{-\lambda_{i1}(t-T)}}{\lambda_{i1}(t-T)} - e^{-\lambda_{i1}(t-T)} \right] \beta_{4t}^i + \left[\frac{1 - e^{-\lambda_{i2}(t-T)}}{\lambda_{i2}(t-T)} - e^{-\lambda_{i2}(t-T)} \right] \beta_{5t}^i - \frac{C_i(t, T)}{T-t}$$

Nesta representação não utilizamos o fator de volatilidade estocástica e mantemos os parâmetros de decaimento λ fixos no tempo em consistência com a especificação afim do modelo, mas estes parâmetros de decaimento são estimados em conjunto com os demais parâmetros do modelo.

6. ESTIMAÇÃO BAYESIANA POR MARKOV CHAIN MONTE CARLO

Em todas as especificações apresentadas até agora temos modelos que podem ser representados com um modelo não-lineares em espaços de estados, onde temos uma equação não-linear de observação para as taxas e um conjunto de equações de estado representando os fatores latentes, e em alguns dos modelos também tratamos os parâmetros de decaimento e a volatilidade como fatores latentes adicionais. Embora a representação básica possa ser estimada através do filtro de Kalman, as formas não lineares não podem ser estimadas por esta metodologia, e mesmo nas representações mais simples existem diversos problemas de estimação neste procedimento.

Dadas as dificuldades computacionais envolvidas na estimação destes modelos, em geral são colocada restrições ad hoc que permitem a estimação, como por exemplo assumir parâmetros de

⁴A expressão analítica para este termo de correção se encontra no apêndice do artigo de [Christensen *et al.*, 2008], e é omitida por questões de espaço.

decaimento fixos e não estimados, ou então realizar a estimação usando procedimentos em dois estágios como comentado na seções 3 e 2.

Neste contexto uma forma de realizar a estimação utilizando toda a informação disponível na estrutura a termo de taxas de juros e evitando a imposição de restrições ad hoc é o uso de métodos Bayesianos de estimação usando algoritmos de Markov Chain Monte Carlo (MCMC). Como mostraremos a seguir, esta metodologia permite tratar os problemas existentes nos mecanismos usuais de estimação, como os problemas de não-linearidade, identificação, e dimensionalidade envolvidos. Na estimação por MCMC modelos lineares e não-lineares são tratados da mesma forma, e uma vantagem da metodologia Bayesiana é que ela permite tratar os fatores latentes como parâmetros adicionais a serem estimados.

Em inferência bayesiana, o objetivo é encontrar a chamada distribuição posterior dos parâmetros de interesse condicionados a amostra observada, denotada por $p(\Theta|y)$. Esta distribuição a posteriori é o resultado da atualização da distribuição a priori assumida para os parâmetros com a informação existente na amostra, representada pela função de verossimilhança.

Para encontrar a distribuição dos parâmetros condicionados a amostra, usa-se a relação derivada do lema de Bayes:

$$(6.1) \quad p(\Theta|y) = p(\Theta, y)/p(y) = p(y|\Theta)p(\Theta)/p(y)$$

onde $p(y|\Theta)$ é a verossimilhança do modelo, $p(\Theta)$ denota a distribuição a priori assumida para o parâmetro e $p(y)$ é a distribuição marginal da amostra, que precisa ser conhecida até uma constante de integração,i.e:

$$(6.2) \quad p(\Theta|y) = p(\Theta, y)/p(\Theta) = p(y|\Theta)p(\Theta)/c$$

e assim a distribuição a posteriori é proporcional ao produto da verossimilhança pela distribuição a priori:

$$(6.3) \quad p(\Theta|y) \propto p(y|\Theta)p(\Theta)$$

Após a obtenção da distribuição a posteriori, a sumarização dos resultados pode ser feita por exemplo calculando-se os valores esperados e a variância da distribuição posteriori de cada parâmetro:

$$(6.4) \quad E(\theta_k|y) = \int \theta_k p(\Theta|y) d\theta$$

$$(6.5) \quad Var(\theta_k|y) = \int \theta_k^2 p(\Theta|y) d\theta - [E(\theta_k|y)]^2$$

e podemos avaliar a densidade marginal de parâmetro θ_j usando:

$$(6.6) \quad p(\theta_j|y) = \int p(\Theta|y) d\theta_1 d\theta_2 \dots d\theta_a$$

Desta forma o objetivo principal da estimação Bayesiana é obter a distribuição a posteriori, que contém a informação a priori atualizada pela informação existente na amostra, dada pela função de verossimilhança. Exceto em alguns casos específicos, em geral utilizando distribuições conjugadas (a distribuição a priori é da mesma família da distribuição a posteriori) , formas analíticas para estas expressões não podem ser obtidas. Mas nestes casos uma forma de se obter estas distribuições, é através de técnicas de integração usando métodos de Monte Carlo. Uma metodologia de Monte Carlo fundamental em métodos de estimação Bayesiana é o uso dos chamados algoritmos de Markov Chain Monte Carlo (MCMC) (e.g. [Robert & Casella, 2005], [Gamerman & Lopes, 2006])

A idéia dos método de MCMC é simular uma cadeia de Markov cuja distribuição estacionária convirja para a distribuição $p(\Theta|y)$. Um resultado fundamental é que a estimação de

$p(\Theta|y)$ pode ser fatorada, utilizando um procedimento de amostragem das distribuições condicionais dos parâmetros, procedimento conhecido como *Componentwise Metropolis-Hastings* (e.g. [Ntzoufras, 2009]). Estas condicionais são de dimensão inferior e que podem ser mais facilmente simuladas. O procedimento pode ser resumido nas seguintes iterações:

$$(6.7) \quad \begin{aligned} & p(\Theta_1|\Theta_2, \Theta_3, \dots, \Theta_n, y) \\ & p(\Theta_2|\Theta_1, \Theta_3, \dots, \Theta_n, y) \\ & \vdots \\ & p(\Theta_n|\Theta_2, \Theta_3, \dots, \Theta_{n-1}, y) \end{aligned}$$

O Teorema de Clifford-Hammersley (veja [Robert & Casella, 2005] para uma derivação deste resultado) assegura que sob certas condições de regularidade este conjunto de distribuições condicionais converge unicamente para a distribuição de $p(\Theta|y)$. Uma vantagem evidente deste método é que ele não envolve nenhuma metodologia de maximização numérica, e desta forma evitamos os problemas numéricos envolvidos em maximização de funções não lineares como as encontradas em nosso problema. A validade empírica desta metodologia é verificada através de métodos para verificar a convergência das cadeias de Markov para a sua distribuição estacionária⁵. Outro ponto importante no uso de métodos de inferência Bayesiana é que o uso de informação a priori ajuda a resolver alguns problemas existentes na estimação clássica, como por exemplo a estimação de modelos não-identificados. Este ponto é discutido em detalhes na seção 8.

Quando todas as distribuições condicionais são conhecidas, o algoritmo de MCMC é o chamado amostrador de Gibbs, aonde a estimação é realizada amostrando-se diretamente das distribuições condicionais. Caso não seja possível amostrar da distribuição condicional analítica, é possível realizar a amostragem destas condicionais utilizando o algoritmo de Metropolis-Hastings, que é uma generalização do método de aceitação-rejeição de simulação de variáveis aleatórias para a amostragem de distribuições condicionais⁶.

Em nosso problema não podemos amostrar diretamente de todas as distribuições condicionais, dada as especificações não-lineares utilizadas e o uso de distribuições não-conjugadas. Desta forma usaremos um algoritmo de Markov Chain Monte Carlo Híbrido, utilizando simultaneamente o algoritmo de Gibbs e algoritmo de Metropolis-Hastings, metodologia proposta inicialmente em [Tierney, 1994]. Neste caso quando temos um condicional conhecida utilizamos Gibbs sampling, e para demais condicionais Metropolis-Hastings. Um algoritmo híbrido de MCMC ([Robert & Casella, 2005]) pode ser visto através como iterações nas seguintes etapas:

$$\begin{aligned} & \text{Para } i=1, \dots, p, \text{ e dados } (\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \dots, \theta_p^{(t)}) \\ & 1 - \text{Simule} \\ & \quad \tilde{\theta}_i \sim q_i(\theta|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \dots, \theta_p^{(t)}) \\ & 2 - \text{Aceite} \end{aligned}$$

$$\theta_i^{(t+1)} = \begin{cases} \theta_i^{(t)} & \text{com probabilidade } 1 - \rho \\ \tilde{\theta}_i & \text{com probabilidade } \rho \end{cases}$$

$$\text{onde } \rho \wedge \begin{cases} \left(\frac{g_{ii}(\tilde{\theta}_i|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \dots, \theta_p^{(t)})}{q_{ii}(\tilde{\theta}_i|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \dots, \theta_p^{(t)})} \right) \\ \left(\frac{g_{ii}(\theta_i^{(t)}|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \dots, \theta_p^{(t)})}{q_{ii}(\theta_i^{(t)}|\theta_1^{(t+1)}, \dots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \dots, \theta_p^{(t)})} \right) \end{cases}$$

onde q é a chamada distribuição tentativa ou auxiliar.

Quando o modelo a ser estimado pode ser colocado em uma formulação de espaço de estado, uma forma conveniente de tratar o problema é através do uso de uma formulação Hierárquica.

⁵Para uma discussão detalhada deste tópico veja [Robert & Casella, 2005], [Gaman & Lopes, 2006] e [Ntzoufras, 2009].

⁶Veja [Robert & Casella, 2005] para uma revisão do amostrador de Gibbs, o algoritmo de Metropolis-Hastings e os demais métodos utilizados a seguir.

Nesta estrutura a representação das priors é baseada em uma hierarquia. Esta formulação é especialmente útil em modelos em espaço de estado⁷, já que a especificação hierárquica permite recuperar a distribuição dos fatores latentes, utilizando como distribuição a priori para o fator latente na data t uma função da posteriori do fator latente no tempo $t-1$. Um exemplo simples é o chamado modelo de nível local::

$$(6.8) \quad \begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \nu_t \end{aligned}$$

Neste exemplo podemos utilizar como distribuição a priori do fator latente μ_t o valor de μ_{t-1} , e assim $\mu_t \sim \pi(\mu_{t-1})$, utilizando diretamente a especificação da equação de estado.

Para caracterizar completamente nosso modelo é necessário discutir quais as distribuições a priori utilizadas. Para os fatores latentes β_{it} e τ_{it} utilizamos como priors o par Normal-Gamma Inversa, através da caracterização hierárquica com média dada pela estrutura de vetor autoregressiva. Para os parâmetros do vetor autoregressivo Φ assumimos uma estrutura normal multivariada com matriz de variância dada por uma distribuição inversa Wishart, e para o fator latente de volatilidade estocástica assumimos $\sigma_t^2 \sim \text{LogNormal}(\phi_0 + \phi_1 \sigma_{t-1}^2, \tau_{\sigma^2})$, com uma distribuição Gamma para τ_{σ^2} , Normal para ϕ_0 e finalmente $\phi_1 \sim \text{Beta}$. Para os demais parâmetros nos processos autoregressivos e na especificação dos parâmetros que identificam os fatores de cada mercado nos modelos de fatores latentes generalizados e globais utilizamos uma estrutura normal-multivariada para a média destes parâmetros e inversa-Wishart para a matriz de variância⁸. Especificações alternativas implementando procedimentos de shrinkage são discutidos na seção 7.

O procedimento de amostragem utiliza o algoritmo de Gibbs para os parâmetros β_{it} , ϕ_0 , parâmetros de processos autoregressivos, parâmetros de cargas no modelo de fator global, parâmetros da distribuição Wishart e as para as distribuições Gamma. Para os demais parâmetros que não possuem distribuição condicional conhecida, ligados às especificações não lineares e não conjugadas, usamos o algoritmo de Metropolis-Hastings e para o parâmetro ϕ_1 nos processos de volatilidade estocástica utilizamos o algoritmo conhecido como *Slice Sampler* ([Neal, 2003]). A especificação do modelo é completa assumindo uma verossimilhança Normal multivariada para a estrutura a termo observada, permitindo recuperar a distribuição a posteriori dos parâmetros através da equação 6.3 com o uso dos algoritmos de estimação por MCMC.

Para obter distribuição preditiva do modelo um passo a frente utilizamos a relação:

$$(6.9) \quad p(\hat{y}_{t+1}|y_t) = \int p(\hat{y}_{t+1}|\Theta)p(\Theta|y_t)p(\Theta)d\Theta$$

que é verossimilhança futura ponderada pela distribuição posterior dos parâmetros, onde y_t são as observações até o período t . Sumarizamos as previsões um passo a frente usando a média e os percentis da distribuição preditiva dada por 6.9.

7. BAYESIAN SHRINKAGE

Os modelos especificados para múltiplas curvas de juros contém um elevado número de parâmetros a serem estimados, principalmente na especificação para a dinâmica dos fatores latentes. A elevada dimensionalidade, em conjunto com os problemas de identificação discutidos na seção 8, tornam a estimação de modelos de estrutura a termo de taxas de juros um problema econométrico bastante complicado.

As soluções usuais para o problema da dimensionalidade do vetor de parâmetros em modelos multimercados envolvem a imposição de restrições ad hoc, que também estão ligados as condições de identificação. No modelo de curva global proposto por [Diebold *et al.*, 2008] uma forma de reduzir o número de parâmetros do modelo é restringir a estimação da equação de observação utilizando apenas os componentes de nível e inclinação, descartando os fatores de curvatura e dupla curvatura. Entretanto este procedimento reduz bastante o ajuste do modelo, em especial

⁷Veja [Koop, 2003] para a estimação de modelos em espaço de estados utilizando a formulação hierárquica.

⁸Para uma discussão sobre estas especificações veja [Bernardo & Smith, 1994].

para as maturidades mais longas. Nestas maturidades estes componentes aumentam de forma significativa o ajuste do modelo (e.g. [Björk & Christensen, 1999], [Christensen *et al.*, 2008]).

No entanto é esperado que muitos destes parâmetros não sejam significantes, e assim possam ser eliminados do modelo. Na especificação do modelo de fatores latentes generalizados, é esperado que os parâmetros de fatores com interpretações distintas não sejam importantes na explicação de outros fatores. Por exemplo é pouco provável que um fator de dupla-curvatura afete os movimentos do fator de inclinação por exemplo. Esta interpretação pode ser justificada pelo desempenho da decomposição de componentes principais de [Litterman & Scheinkman, 1991], aonde os componentes são por construção ortogonais.

Na estimação Bayesiana este problema é implicitamente tratado através da estrutura de prioris utilizada. Um parâmetro que tenha valor posterior esperado zero é normalmente especificado através de uma distribuição a priori concentrada no valor zero. Esta é a interpretação da chamada ‘Minnesota’ prior (e.g. [Doan *et al.*, 1984]) utilizada em modelos de séries temporais, que advoga uma priori para séries temporais não estacionárias definindo um processo de passeio aleatório em modelos de vetores autoregressivos, impondo uma priori centrada em zero para os defasagens superiores a um e centrada em um para a primeira defasagem da variável

Uma forma alternativa de tratar este problema é através do uso de técnicas conhecidas como *Bayesian Shrinkage*, que é o uso de prioris que coloquem um peso maior no zero do que as prioris padrões. Neste artigo utilizamos duas formas de priors de *shrinkage*. A primeira forma é o uso da prior Laplace (Exponencial Dupla) e a segunda forma é o uso da prior ‘Minnesota’ generalizada.

A estimação utilizando a priori Laplace (Exponencial Dupla) está relacionada com o método de estimação conhecido como *LASSO - Least Absolute shrinkage and selection operator*, proposto por [Tibshirani, 1996]. O estimador LASSO é obtido como a solução de um problema de estimação utilizando uma penalidade ℓ_1 na minimização do problema:

$$(7.1) \quad ||Y - X\beta|| + \lambda \sum_{j=2}^q |\beta_j|$$

Uma vantagem do estimador LASSO é que ao invés de apenas forçar os estimadores para zero, como no caso de técnicas como ridge regression, ele efetivamente permite que alguns estimadores sejam identicamente iguais a zero, realizando simultaneamente um procedimento de *shrinkage* e seleção de modelos. O estimador LASSO pode ser interpretado como uma estimação de moda a posteriori em um contexto Bayesiano, através de uma distribuição a priori Laplace (Exponencial Dupla), como apontam o próprio [Tibshirani, 1996] e [Park & Casella, 2008]. Esta distribuição Laplace é uma função de dois hiperparâmetros (μ, b) na forma:

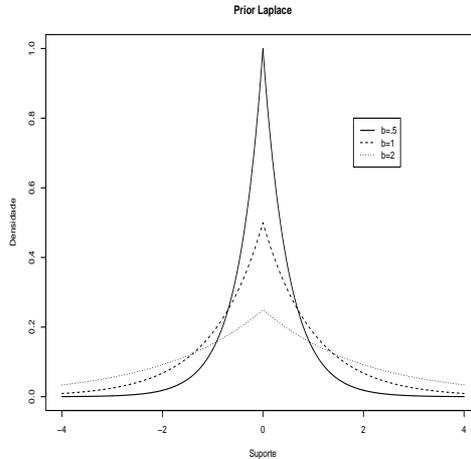
$$(7.2) \quad \pi(\beta) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}},$$

onde (μ, b) podem ser interpretados como fatores de locação e escala. A figura 7.1 mostra diferentes especificações da priori Laplace, mostrando que o peso em zero é muito maior do que por exemplo prioris Gaussianas com o mesmo fator de escala. A única dificuldade associada com o uso de prioris Laplace é que esta distribuição não é conjugada, e assim é necessária uma etapa adicional no procedimento de estimação utilizando o algoritmo de Metropolis-Hastings. Adotamos esta estrutura de prioris Laplace independentes para os parâmetros nos vetores autoregressivos que definem os fatores latentes, e utilizamos como valores de μ e b os valores 0 e .1.

A prior de ‘Minnesota’ generalizada proposta em [Robertson, 1999] e [Kadiyala & Karlsson, 2007], e advogada para o uso na estimação de modelos de vetores autoregressivos Bayesianos de elevada dimensão por [Banbura *et al.*, n.d.], consiste em uma generalização de priori proposta por [Doan *et al.*, 1984]. Nesta formulação temos que a priori para as matrizes de parâmetros Φ_i e Φ_j nos modelos de fatores generalizados é dada por:

$$(7.3) \quad \text{vec}[\Phi_i] = N \left(\text{vec}(\Phi_{i0}), \psi \otimes \Omega_0 \right)$$

FIGURA 7.1. Laplace Prior



onde vec é operador que empilha os parâmetros em uma coluna e

$$(7.4) \quad \psi \sim iW(S_0, \alpha_0)$$

onde $E(S_0) = \begin{cases} \delta_i \text{ se } i = k \\ 0 \text{ c.c.} \end{cases}$ e $\alpha_0 = \begin{cases} \lambda^2 \text{ se } i = k \\ \lambda^2 \frac{\sigma_i}{\sigma_i} \text{ c.c.} \end{cases}$ e iW denota uma distribuição Wishart

Inversa.

Neste caso assumimos δ_i é a variância esperada de cada fator latente e o parâmetro λ controla o fator de *shrinkage* escolhido e σ_i e σ_j são as variabilidades dos fatores latentes. Para os modelos tratados neste artigo utilizamos um fator de shrinkage $\lambda=.1$.

8. IDENTIFICAÇÃO

Note que na definição das possíveis especificações possíveis para os modelos dinâmicos para a estrutura a termo de taxas de juros, existe um trade-off entre uma especificação mais rica e a dificuldade de estimação computacional. [Dai & Singleton, 2000] apontam que os problemas na especificação de modelos afins de estrutura a termo de taxas de juros envolvem condições de admissibilidade, i.e. o modelo leva a preços de títulos bem definidos, e condições de identificação econométrica.

O conceito de identificação em modelos econométricos pode ser resumido como a propriedade de que um modelo é considerado não-identificado se existe mais de um vetor de parâmetros que define uma função de verossimilhança equivalente, e esta perspectiva é válida tanto em modelos clássicos (e.g. [Rothemberg, 1971]) como em modelos Bayesianos (e.g. [Kadane, 1974], [Poirier, 1998] e [Aldrich, 2002]). Formalmente temos que um modelo é não identificado se considerando uma função de verossimilhança regular $L(\theta; y)$, onde $\theta \in \Theta$ é um vetor $K \times 1$ e $\Theta \in R^K$. Se θ é não identificado para todo $\theta^{(1)} \in \Theta$ existe um outro $\theta^{(2)} \in \Theta$ tal que $L(\theta^{(1)}; y) = L(\theta^{(2)}; y)$ para todo y . Desta forma em modelos não-identificados existe mais de um vetor satisfazendo a função critério de estimação, o máximo da função de verossimilhança.

Conforme [Kadane, 1974], identificação é uma propriedade da função de verossimilhança, e assim a identificação é a mesma considerando perspectivas clássicas ou Bayesianas. Como coloca [Poirier, 1998] a solução para a estimação de modelos não-identificados é a mesma sob as duas perspectivas, através do uso de mais informação no modelo, e esta informação em geral não está contida na amostra. A solução proposta para problemas de identificação na perspectiva clássica em geral está representada na imposição de restrições no espaço paramétrico, usualmente eliminando parâmetros redundantes no modelo. A perspectiva Bayesiana é menos dogmática, e normalmente identificação é obtida através do uso de informação a priori. Citando [Poirier, 1998]:

“A Bayesian analysis of a nonidentified model is always possible if a proper prior on all the parameters is specified”

A estimação de modelos não-identificados através da priori é obtida notando que o uso de uma priori adequada pode reduzir o espaço amostral possível na distribuição posterior dos dados, assim reduzindo a probabilidade da posterior se situar em uma região de não-identificação. Como colocado formalmente em [Florens *et al.*, 1990] a escolha de uma priori adequada possibilita a estimação de uma posteriori identificada através da redução da sigma-álgebra geradora da distribuição a posteriori dos parâmetros de interesse. Entretanto algumas considerações são necessárias. A primeira é que podem existir situações aonde os dados são não informativos, e as distribuições a priori e a posteriori são equivalentes, e também podem existir situações problemáticas se são utilizadas prioris impróprias. Uma discussão detalhada destes problemas podem ser encontrada em [Poirier, 1998]. Esta mesma referência contém a discussão de duas situações diretamente relacionadas a estimação no contexto do problema da estimação dos modelos de estrutura a termo propostos neste artigo.

A primeira discussão está relacionada ao problema de multicolinearidade. Note que na forma geral dos modelos de Nelson-Siegel-Svensson (eq. 4.1) existe um problema potencial de multicolinearidade que leva a não-identificação do modelo. Os termos relacionados aos componentes β_2 e β_4 são potencialmente não identificados para valores próximos destes dois fatores, o que inclui a especificação do valor dos parâmetros de decaimento. Normalmente é assumido que $\tau_2 > \tau_1$ para a identificação do modelo, como em [Christensen *et al.*, 2008]. No caso de multicolinearidade identificação pode ser obtida através de uma priori adequada para os parâmetros relevantes, e no caso da estimação dos parâmetros de decaimento a identificação é obtida assumindo-se prioris que conduzam a posteriori destes parâmetros a distribuições aonde a probabilidade posterior de se observar $\tau_2 > \tau_1$ seja máxima.

A segunda discussão relevante está relacionada a estimação de modelos hierárquicos, aonde é possível mostrar que identificação pode ser obtida através de uma escolha adequada para a distribuição a priori nos parâmetros envolvidos em cada hierarquia do modelo. Na formulação proposta para o modelo, utilizamos uma representação em espaço de estado para a evolução dos fatores latentes do modelo, e a estimação deste representação é dada através da formulação hierárquica aonde a priori para o fator latente em $t+1$ é a posteriori estimada para este fator latente no período t . Neste caso temos que a condição de informatividade dos dados é sempre respeitada, e com uma escolha adequada de prioris o modelo pode sempre ser estimado, evitando os problemas de não-identificação normalmente existentes na estimação clássica utilizando o filtro de Kalman (e.g. [Duffe, 2002]). Este problema e suas implicações econômicas é discutido em [Kim & Orphanides, 2005], que aponta que modelos afins de estrutura a termo podem ser caracterizados por estimações observacionalmente equivalentes mas com interpretações econômicas bastante distintas.

Note que uma estrutura a termo de taxas de juros contém estatísticas suficientes para identificar os parâmetros necessários, dadas pelas observações passadas da curva de juros. Este é exatamente um dos problemas da estimação em dois estágios utilizada em [Diebold & Li, 2006, Diebold *et al.*, 2008], que é o fato da estimação do primeiro estágio justamente ignorar toda a estrutura de dependência temporal entre os fatores latentes. A estimação simultânea Bayesiana utiliza o parâmetro estimado no período $t-1$ como a priori para o parâmetro em t , e como em geral esta estimação é informativa, conseguimos a redução discutida em [Florens *et al.*, 1990] na sigma-álgebra geradora da distribuição posterior do fator latente, resolvendo o problema de identificação para os modelos em questão.

9. APLICAÇÃO EMPÍRICA

9.1. Base de dados. Para realizar a análise empírica dos modelos propostos utilizamos curvas de juros de dois mercados distintos. A primeira curva é construída utilizando-se dados da estrutura a termo da curva de Cupom Cambial. O Cupom Cambial pode ser resumido como uma estrutura a termo de instrumentos negociados no Brasil, mas com rendimentos em dólares. Outros estudos para a modelagem da curva de cupom Cambial são [Pinheiro *et al.*, 2007] que modela esta curva

TABELA 1. Estatísticas Descritivas

Cupom Cambial	6	9	12	24	36	48	60	72	84	96	108	120
Min.	0.0292	0.0309	0.0323	0.0365	0.0401	0.0415	0.0428	0.0441	0.0457	0.0476	0.0495	0.0515
1st Qu.	0.0410	0.0436	0.0449	0.0483	0.0523	0.0554	0.0582	0.0610	0.0640	0.0671	0.0704	0.0735
Mediana	0.0574	0.0569	0.0564	0.0557	0.0572	0.0603	0.0634	0.0669	0.0705	0.0744	0.0779	0.0809
Média	0.0532	0.0537	0.0540	0.0547	0.0567	0.0595	0.0628	0.0668	0.0708	0.0749	0.0790	0.0831
3rd Qu.	0.0631	0.0625	0.0623	0.0615	0.0622	0.0646	0.0680	0.0724	0.0766	0.0811	0.0858	0.0906
Max.	0.0942	0.0894	0.0826	0.0823	0.0843	0.0867	0.0913	0.1021	0.1125	0.1224	0.1323	0.1425
desv. pad	0.0124	0.0116	0.0110	0.0092	0.0082	0.0079	0.0083	0.0094	0.0107	0.0121	0.0136	0.0153
Eurodólar	6	9	12	24	36	48	60	72	84	96	108	120
Min.	0.0183	0.0185	0.0187	0.0205	0.0233	0.0263	0.0286	0.0305	0.0320	0.0331	0.0340	0.0347
1st Qu.	0.0281	0.0281	0.0281	0.0304	0.0332	0.0358	0.0381	0.0401	0.0419	0.0434	0.0447	0.0458
Mediana	0.0361	0.0341	0.0333	0.0358	0.0387	0.0408	0.0425	0.0440	0.0454	0.0466	0.0478	0.0488
Média	0.0389	0.0382	0.0377	0.0381	0.0400	0.0418	0.0434	0.0448	0.0461	0.0473	0.0483	0.0492
3rd Qu.	0.0526	0.0510	0.0499	0.0482	0.0481	0.0485	0.0491	0.0498	0.0504	0.0511	0.0516	0.0523
Max.	0.0538	0.0537	0.0538	0.0540	0.0546	0.0553	0.0560	0.0566	0.0572	0.0578	0.0584	0.0589
desv. pad.	0.0119	0.0116	0.0112	0.0097	0.0083	0.0073	0.0065	0.0058	0.0053	0.0049	0.0046	0.0044

usando uma estrutura polinomial com variáveis latentes e [Pereira, 2009] que utiliza uma forma simplificada do modelo de [Diebold & Li, 2006].

Esta curva foi construída através de um instrumento sintético calculado com base em ativos transacionados na BM&F. O cupom cambial foi calculado por não-arbitragem igualando-se o retorno do DDI, que é um instrumento de renda fixa cuja remuneração é acrescida com a rentabilidade acumulada do CDI. A fórmula utilizada para o cálculo do Cupom Cambial é dada por:

$$(9.1) \quad C_t = \left(\frac{\prod_{t=1}^T (1 + i_t)^{\left(\frac{1}{252}\right)}}{(1 + \Delta e_t)} \right) * \frac{360}{T}$$

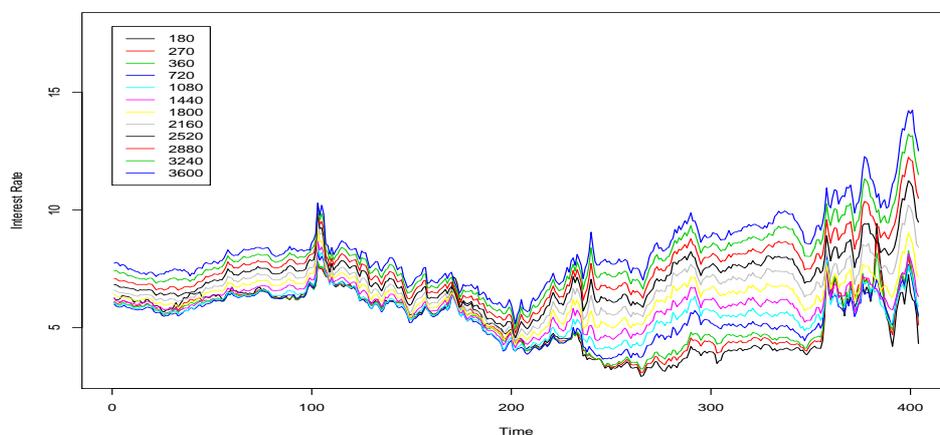
onde T é o número de dias corridos entre a negociação do contrato e o vencimento, i_t é a taxa CDI negociada no mercado interbancário no dia t e Δe_t é a valorização cambial medida pela taxa de câmbio em reais por dólar (PTAX800) observada entre o dia útil anterior a data de operação no mercado futuro e o último dia do mês anterior ao vencimento do contrato. Como neste contrato existe uma distorção dada pela utilização da PTAX do dia anterior, utilizamos a metodologia de replicação deste título através de instrumentos de mercado com mais liquidez, usando o dólar spot, dólar futuro, futuros de DI e Forward Rate Agreements. Esta metodologia foi utilizada em [Pereira, 2009], que discute em detalhes as vantagens deste procedimento.

A outra curva utilizada neste estudo é uma curva de juros construída com base nas remunerações obtidas no mercado de Eurodólar, que corresponde ao mercado de depósitos financeiros em dólares negociado fora dos EUA. Esta curva externa foi construída utilizando-se os contratos a termo de Eurodólar negociados na Chicago Mercantile Exchange. Estas curvas são construídas utilizando-se a metodologia sugerida por [Burghardt, 2003]. Note que estes dois instrumentos são escolhidos de forma a terem uma remuneração em uma mesma moeda, e assim eliminar a influência da variação cambial nos retornos em diferentes mercado.

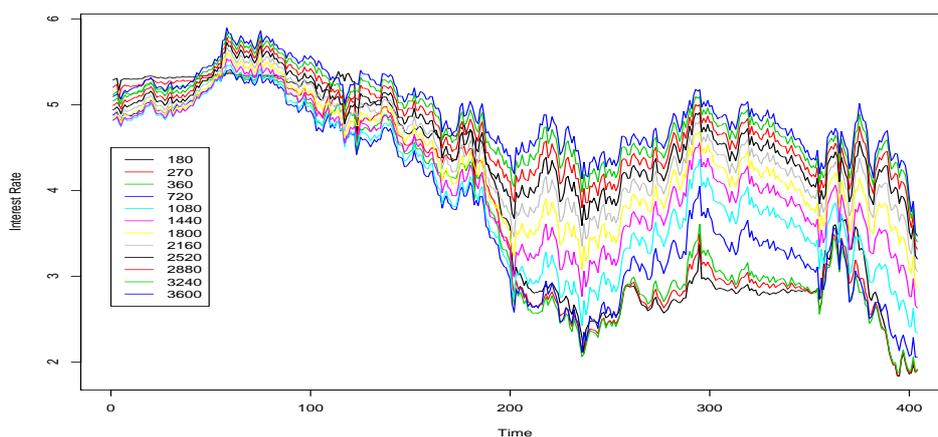
Para as duas curvas trabalhamos com vértices fixos de 6, 9, 12, 24, 36, 48, 60, 72, 84, 96, 108 e 120 meses, com a amostra indo de 6/3/2007 até 26/11/2008, contendo 402 observações. As estatísticas descritivas para cada vértice destas duas curvas está colocado na Tabela 1, e a Figura 9.1 mostra a evolução destas duas curvas no tempo.

É possível observar algumas características nestas duas curvas. Como aponta [Pereira, 2009] nas duas curvas existe um movimento de aumento na volatilidade média e também nos spreads. O período final da curva de juros de Eurodólar reflete as cortes nas taxas de juros realizadas pelo Federal Reserve.

FIGURA 9.1. Curvas de Juros



(a) Curva Cupom Cambial



(b) Curva Eurodólar

Estes padrões nas curvas de juros as tornam interessantes como objeto de estudo para os modelos propostos. É possível notar que existe uma grande variabilidade no formato das curvas de juros no tempo, indicando que os fatores latentes devem ter grande variabilidade nestas duas curvas. Outro ponto é que fica evidente que o padrão de decaimento da curva se modifica bastante no tempo, justificando o uso de parâmetros variantes no tempo em oposição a fixação realizada na estimação usual do modelo de [Diebold *et al.*, 2005]. Outro ponto interessante é dado pela estrutura de volatilidade, que não é constante no tempo e justifica o uso do componente de volatilidade estocástica.

9.2. Análise Comparativa. Para realizar uma análise completa das três classes de modelos propostos, realizamos a estimação dos modelos completos de cada classe e também com a imposição de sub-modelos de cada classe. Estas diferentes especificações permitem analisar como as diferentes características dos modelos afetam o ajuste do modelo e os resultados obtidos. Foram estimados 13 especificações diferentes, detalhadas abaixo:

- (1) Curvas Independentes - nesta especificação as curvas são independentes - os fatores latentes de cada curva só dependem dos demais fatores da mesma curva, ignorando a interdependência com o outro mercado. Esta especificação corresponde ao modelo de fatores latentes generalizados, com a restrição que os parâmetros Φ , θ e γ correspondentes a curva do outro mercado são eliminados da especificação das equações 4.6, 4.7 e 4.8.
- (2) Modelo Fator Latente Generalizado Completo - este modelo corresponde as equações 4.5, 4.6, 4.7 e 4.8 com todos os parâmetros sendo estimados.
- (3) Modelo Fator Latente Generalizado, mas fator cruzado restrito - Neste modelo assumimos que a matriz Φ na equação 4.6 tem posto completo para os fatores da mesma curva e é uma matriz diagonal para os fatores da outra curva de juros.
- (4) Modelo Fator Latente Generalizado Diagonal - nesta especificação assumimos que as matrizes Φ_i e Φ_j são diagonais, e assim cada fator só depende dele em $t-1$ e do fator equivalente da outra curva em $t-1$. Por exemplo o nível em t da curva de cupom cambial só depende do nível da curva de cupom cambial em $t-1$ e do nível da curva de Eurodólar em $t-1$ e não depende dos demais fatores.
- (5) Modelo Fator Latente Generalizado Triangular - curva de cupom cambial depende dela mesma em $t-1$ e da curva de Eurodólares, mas a curva de Eurodólares só depende dela própria - Nesse modelo assumimos que a curva local é influenciada pela curva estrangeira, mas a curva estrangeira é independente da curva de Cupom Cambial, assumindo uma estrutura triangular.
- (6) Modelo de Fator Global Generalizado Identificado com Curva Eurodólar. Neste modelo assumimos que o fator global é dado pelos fatores latentes da curva de Eurodólar. Desta forma neste modelo a curva de Cupom Cambial é um deslocamento direto da curva de Eurodólar mais um fator idiossincrático. Note que esta estrutura é muito mais simples que o fator global generalizado completo, já que estimamos diretamente os fatores latentes de Eurodólar e obtemos os fatores da curva de cupom cambial estimando apenas os loadings correspondentes.
- (7) Modelo de Fator Global Completo - Nesta especificação assumimos a estrutura completa de fator global generalizado, aonde tanto os fatores da curva de Eurodólar e da curva de Cupom Cambial são deslocamentos da fator global latente mais fatores idiossincráticos, correspondendo as equações 4.9- 4.15 do modelo de fatores globais generalizado.
- (8) Modelo de Fatores Latentes Generalizado com Bayesian Shrinkage via Prior Laplace - Nesta especificação estimamos o modelo completo de fatores latentes generalizado, mas utilizando a estrutura de Prioris Laplace (equações 7.1, 7.2) para os parâmetros autoregressivos do modelo. Neste caso assumimos que (μ, b) são dados pelo vetor $(0, 1)$.
- (9) Modelo de Fatores Latentes Generalizado com Bayesian Shrinkage via *Generalized Minnesota Prior* - Novamente o modelo de fatores latentes generalizado completo, mas agora utilizando a estrutura da Priori Minnesota Generalizada descrita pelas equações 7.3e 7.4.
- (10) Modelo Generalizado com 5 fatores - Utilizamos a especificação de 5 fatores dada pela equação 5.12, mas sem a correção para não-arbitragem e assumindo a estrutura de fator latente generalizado com interações entre curvas de juros. Este modelo é análogo ao proposto por [Björk & Christensen, 1999]. Assumimos nesta especificação que os parâmetros de decaimento e volatilidade estocástica são constantes. Este modelo tem o objetivo de verificar o ganho dado por estes dois fatores adicionais no ajuste das curvas de juros.
- (11) Modelo Generalizado com 5 fatores e condições de Não-Arbitragem - Utilizamos a especificação completa do modelo de [Christensen *et al.*, 2008], mas assumindo a estrutura de fatores latentes generalizados permitindo a interação entre fatores latentes de diferentes curvas e utilizando a correção por Não-Arbitragem dada pelas equações 5.13e 5.15. Este modelo generaliza o modelo de [Christensen *et al.*, 2008] para mais de um mercado.
- (12) Modelo Generalizado com 5 fatores e Bayesian Shrinkage. Este modelo é similar ao modelo 10, mais utilizando a estrutura de shrinkage via prior Laplace.
- (13) Modelo Generalizado com 5 fatores, Não-Arbitragem e Bayesian Shrinkage. Esta especificação corresponde ao modelo 11 com a correção de Não-Arbitragem mais o uso de shrinkage via prior Laplace.

TABELA 2. DIC

Modelo	DIC
Modelo 1	-41240.46
Modelo 2	-90263.49
Modelo 3	-42076.37
Modelo 4	-73328.45
Modelo 5	-96747.54
Modelo 6	-5774.245
Modelo 7	-115687.7
Modelo 8	-115814.7
Modelo 9	-74041.7
Modelo 10	-97506.19
Modelo 11	-97506.19
Modelo 12	-101149.4
Modelo 13	-101149.4

A estimação de todos estes modelos utiliza um período de burn-in (número de amostras descartadas) de 5000 iterações, e mais 10000 iterações para a construção das distribuições posteriores. A verificação de convergência das cadeias foi realizada utilizando os procedimentos de Geweke e Gelman-Rubin (e.g. [Ntzoufras, 2009]) e indicaram que não houveram problemas na convergência das cadeias de Markov simuladas.

O primeiro mecanismo de comparação entre modelos usa o critério de informação DIC (Deviance Information Criteria) de [Spiegelhalter *et al.*, 2002]. O DIC é um critério de informação Bayesiano que permite realizar seleção de modelos de forma análoga aos critérios BIC e AIC normalmente utilizados. Este critério é interessante na comparação de modelos complexos com elevado número de parâmetros, já que no DIC a penalização é sobre o número efetivo de parâmetros conforme definido em [Spiegelhalter *et al.*, 2002]. O DIC também tem a característica de obter resultados equivalentes a versão robusta do critério AIC (e.g. [Claeskens & Hjort, 2008]) e assim também é válido como um critério de seleção em uma perspectiva de inferência clássica.

A Tabela 2 mostra o DIC estimado para os modelos estimados. Por este critério os melhores modelos são os modelos 7 e 8, que correspondem ao Modelo de Fator Global Generalizado completo e ao Modelo de Fatores Latentes Generalizados utilizando shrinkage através da Priori Laplace. O modelo de fator global possui um número de parâmetros bastante inferior ao modelo de fatores generalizados, mas possui uma estrutura mais complexa dado aos fatores latentes globais. O fato de o DIC destes dois modelos serem equivalente mostram que o ajuste dentro da amostra destes dois modelos é equivalente utilizando a penalização pelo número efetivo de parâmetros.

Também é importante notar que o uso da Prior Laplace permite reduzir de forma significativa a complexidade do modelo, já que comparando o DIC do modelo 2 com o modelo 8 notamos uma redução bastante significativa. Também é interessante notar que o pior modelo pelo DIC foi o modelo de fatores independentes, mostrando que a interação entre as curvas adiciona poder de ajuste nestes modelos de estrutura a termo de taxas de juros, mesmo penalizando para a maior complexidade do modelo.

O resultado obtido pelo DIC aponta que a estrutura mais geral dos modelos 7 e 8 efetivamente aumenta o ajuste dentro da amostra, mesmo penalizando pela maior complexidade destes modelos. Outro ponto importante é que o DIC corresponde ao ajuste global do modelo, e assim não permite diferenciar o ajuste relativo para cada curva de juros ou para cada vértice em particular. Mas uma questão relevante é verificar se esta maior complexidade existente nos modelos selecionados pelo critério DIC leva a um melhor o poder de previsão.

Para realizar esta análise realizamos uma análise de poder preditivo, comparando os modelos estimados através de vários critérios de análise de previsão. Calculamos para previsões um passo-a-frente os critérios de previsão ME (*Mean Error*), RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*), MPE (*Mean Percentage Error*), MAPE (*Mean Absolute Percentage Error*) e o

TABELA 3. Previsões um passo a frente - Curva Cupom Cambial

	ME	RMSE	MAE	MPE	MAPE	Theil's U
Modelo 1	4.948467e-05	0.0033243248	0.0023628392	-0.10506614	3.6941737	1.3876540
Modelo 2	3.120740e-05	0.0061732919	0.0046305395	-0.67884277	7.2658601	2.5536768
Modelo 3	3.120740e-05	0.0061732919	0.0046305395	-0.67884277	7.2658601	2.5536768
Modelo 4	-8.001256e-06	0.0021101813	0.0014223738	-0.09677698	2.2386319	0.9026644
Modelo 5	2.929328e-06	0.0061070020	0.0045121777	-0.71735752	7.0040327	2.4440608
Modelo 6	-1.134294e-04	0.0065386572	0.0021641319	-0.38978741	3.4117078	2.4189061
Modelo 7	-9.932863e-05	0.0041431653	0.0008865839	-0.17691421	1.4313262	1.5970411
Modelo 8	1.077975e-07	0.0046493827	0.0032573506	-0.37735241	5.0503655	1.8872831
Modelo 9	3.035743e-06	0.0054047840	0.0041093530	-0.54053051	6.5118160	2.2959305
Modelo 10	-1.568414e-06	0.0014702792	0.0011108274	-0.02515509	1.7473377	0.6167394
Modelo 11	7.119768e-05	0.0014720012	0.0011046995	0.09479088	1.7361130	0.6173000
Modelo 12	-7.629270e-07*	0.0007277954*	0.0005350270*	-0.00980893*	0.8361936*	0.3065305*
Modelo 13	2.347561e-04	0.0007647195	0.0005653953	0.37841530	0.8928385	0.3248396
*denota	o melhor	modelo				

critério *Theil's U*. Propriedades destas medidas de acurácia de previsão podem ser encontradas em [Hyndman & Koehler, 2006]. As Tabelas 3 e 4 mostram estas medidas para as curvas de Cupom Cambial e Eurodólar⁹.

Os resultados obtidos para a curva de Cupom Cambial mostram que o melhor modelo em todos os critérios de previsão é o modelo 12, que é o modelo de 5 fatores, obtendo um poder de previsão bastante superior aos demais modelos. Podemos interpretar este resultado como o fato de que a adição de dois fatores adicionais de inclinação e curvatura ajuda a um melhor ajuste e previsão nesta curva de juros, o que era esperado pela maior variação nos formatos da curva de Cupom Cambial no tempo. Esta conclusão pode ser observada através do critério *Theil's U*, que mostra o ganho relativo de previsão em relação a uma previsão ingênua usando um passeio aleatório.

Podemos observar que sistematicamente temos que os modelos com 5 fatores, com e sem correção para Não-Arbitragem, conseguem obter um desempenho superior a previsão de passeio aleatório na curva de Cupom Cambial. Embora seja uma amostra diferente, é possível comparar os resultados com os obtidos com os modelos para a curva de cupom cambial utilizados em [Pinheiro *et al.*, 2007]. Para a previsão um dia a frente, o menor *Theil's U* obtido nesse estudo é de .88, enquanto que conseguimos uma redução para 0.3065 utilizando o modelo 12, mas notando que as amostras são de períodos diferentes e os vértices estudados também são diferentes e assim esta comparação é informal.

Outro comentário importante é que a correção para não-arbitragem não reduz de forma significativa o poder preditivo dos modelos, o que podemos observar comparando os modelos 10-11 e 12-13. No critério de erro percentual absoluto médio (MAPE) vemos que a correção para não-arbitragem melhora o poder preditivo do modelo quando comparamos o modelo 10 com o modelo 11, e assim a correção por não-arbitragem se colocada em um modelo com flexibilidade suficiente, como o modelo de 5 fatores, não representa uma perda de poder preditivo. Desta forma conseguimos alcançar o melhor de dois mundos - não-arbitragem e acurácia nas previsões.

Para a curva de Eurodólar os resultados gerais mostram que todos os modelos tem um desempenho preditivo adequado, o que podemos observar pelo critério de *Theil's U* que mostra que todos os modelos tem um desempenho muito superior a previsão ingênua usando o passeio aleatório, ou então pelo critério MAPE que mostra o baixo erro percentual absoluto médio. Neste caso é interessante notar que pelos critérios MAE e MAPE o melhor modelo é o modelo 7, que é o modelo de fator global generalizado completo e pelos demais critérios temos que o melhor modelo é o modelo independente.

⁹Esta curvas correspondem a uma agregação dos erros de previsão para todos os vértices das curvas de juros, mas também foram calculados medidas para cada vértice em separado. Estes resultados não são mostrados por questões de espaço, mas são disponíveis com os autores.

TABELA 4. Previsões - Curva Eurodólar

	ME	RMSE	MAE	MPE	MAPE	Theil's U
Modelo 1	-6.320870e-08*	0.0001879464*	0.0001483558	-0.002479230*	0.3757145	0.08551584*
Modelo 2	-1.202949e-06	0.0007612598	0.0006331330	-0.019946197	1.6084661	0.35552141
Modelo 3	-6.134617e-07	0.0007175650	0.0005976160	-0.016846364	1.5134021	0.33254931
Modelo 4	-2.514408e-06	0.0003228174	0.0002566345	-0.010678224	0.6317814	0.14342737
Modelo 5	-1.030380e-05	0.0002885804	0.0002268257	-0.032457901	0.5619791	0.13099188
Modelo 6	1.213102e-05	0.0003598094	0.0002831787	0.021745183	0.7176587	0.16520193
Modelo 7	8.326324e-06	0.0002182445	0.0001327137*	0.014453269	0.3432117*	0.11512664
Modelo 8	1.603705e-07	0.0002768067	0.0002224759	-0.002767453	0.5307122	0.11509083
Modelo 9	-1.571455e-06	0.0005298119	0.0004375458	-0.013132845	1.0805472	0.23450715
Modelo 10	-3.153238e-07	0.0009584696	0.0007179616	-0.052340573	1.8029979	0.42355519
Modelo 11	2.786841e-05	0.0009588747	0.0007156113	0.017230232	1.7958090	0.42326406
Modelo 12	-1.035208e-06	0.0013243624	0.0010372927	-0.093306970	2.6140207	0.59081940
Modelo 13	1.093918e-04	0.0013288721	0.0010340418	0.179279071	2.6029361	0.59052908

*denota o melhor modelo

Estes resultados podem ser interpretados notando que a curva de Eurodólar deve ser a priori muito menos sensível a influências das demais curvas de juros, e desta forma o melhor resultado preditivo do modelo de curvas independentes faz sentido. Mudanças na curva de juros de Cupom Cambial não devem ter poder preditivo significativo sobre a curva de Eurodólar, e dada a menor complexidade deste modelo, esta característica se reflete em um menor erro médio de previsão. No entanto podemos notar que em geral os modelos se caracterizam por um viés negativo na previsão, o que não acontece no modelo 7 que é o melhor modelo pelos critérios MAE e MAPE. O resultado obtido pelo modelo 7 pode ser explicado pelo formato dos fatores globais estimados, que são mais próximos da curva de Eurodólar do que da Curva de Cupom Cambial, como é mostrado na Figura 9.3.

No caso da curva de Eurodólar a adição dos fatores adicionais de inclinação e curvatura não representa um melhor poder preditivo, e os modelos 10-13 tem um desempenho bastante inferior aos demais modelos. Este resultado é consistente com o fato estilizado que o formato da curva de juros de países desenvolvidos é mais simples que curvas de países emergentes, e assim o poder preditivo dos modelos mais simples é maior para esta curva do que para a curva de Cupom Cambial, que necessita de uma especificação mais flexível.

9.3. Importância da Correção de Não-Arbitragem. Embora a consistência com não-arbitragem seja uma condição fundamental na especificação de modelos de estrutura a termo de taxas de juros, uma questão interessante é verificar se as curvas de juros observadas são consistentes ou não com não-arbitragem. No contexto da formulação de [Christensen *et al.*, 2008] é possível medir este efeito olhando para o fator de correção $C_i(t, T)$ na equação 5.15. Note que este modelo é basicamente uma curva de juros baseada no modelo de [Björk & Christensen, 1999] mais este fator que assegura a consistência com não-arbitragem. Se a magnitude deste fator é muito baixa e não-significante temos uma evidência que a própria curva ajustada já é livre de arbitragem e este fator de correção não é necessário. Uma forma de verificar este efeito é olhando para a distribuição a posteriori do fator de correção por não arbitragem estimada pelo algoritmo de Markov Chain Monte Carlo.

A Tabela 5 mostra os percentis de 2.5%, 50% e 97.5% para a distribuição a posteriori estimada para o fator de correção por não arbitragem dado pela termo $C_i(t, T)$ na equação 5.15, para cada maturidade estimada nas curvas de Cupom Cambial e Eurodólar. Estes percentis estimados podem ser interpretados como um intervalo de credibilidade para o fator de correção por não-arbitragem, e assim podem ser interpretados em uma perspectiva frequentista como uma estimação por intervalos e também como um conseqüente teste de hipóteses via inversão da estimação por intervalos (e.g. [Bernardo & Smith, 1994]). A hipótese nula de interesse seria que a correção por não-arbitragem em cada maturidade é igual a zero, contra uma hipótese alternativa de que esta correção é diferente

TABELA 5. Distribuição Posterior - Fator de Correção por Não-Arbitragem

Cupom Cambial	quantile 2.5 %	quantile 50 %	quantile 97.5 %	Eurodólar	quantile 2.5 %	quantile 50 %	quantile 97.5 %
6	-2.505592e-04	5.627831e-05	0.0003210676	6	-2.923742e-04	-1.010111e-04	1.133912e-05
9	-2.220485e-04	5.973929e-05	0.0003037833	9	-2.099763e-04	-6.273529e-05	1.253434e-05
12	-1.978511e-04	6.004418e-05	0.0002859084	12	-1.570117e-04	-3.717365e-05	2.503620e-05
24	-1.148664e-04	5.428410e-05	0.0002036099	24	-2.398569e-05	2.081719e-05	5.001232e-05
36	-3.272735e-05	5.953053e-05	0.0001443693	36	-1.781659e-05	5.687899e-05	9.922963e-05
48	4.570273e-05	7.687623e-05	0.0001133094	48	-1.364977e-05	8.434156e-05	1.812485e-04
60	2.077555e-05	1.017708e-04	0.0001943380	60	-7.023459e-05	1.013277e-04	2.686080e-04
72	-3.571103e-05	1.293691e-04	0.0003155333	72	-1.510072e-04	1.114461e-04	3.622645e-04
84	-1.008644e-04	1.601904e-04	0.0004644550	84	-2.554487e-04	1.097855e-04	4.644888e-04
96	-1.747916e-04	1.919751e-04	0.0006244936	96	-3.797314e-04	1.102472e-04	5.769284e-04
108	-2.651789e-04	2.223300e-04	0.0007991271	108	-5.257737e-04	9.985080e-05	6.997016e-04
120	-3.731035e-04	2.527459e-04	0.0009915839	120	-6.933878e-04	8.184704e-05	8.314890e-04

de zero. Neste caso a validade desta hipótese nula pode ser testada verificando se os intervalos de credibilidade obtidos da distribuição posterior do fator de correção por não-arbitragem incluem o valor pontual zero ou não.

A Tabela 5 mostra que em geral os valores estimados para o fator de correção por não-arbitragem tem magnitude reduzida, e que apenas as maturidades de 1440 e 1800 dias para a curva de Cupom Cambial não incluem o zero dentro dos intervalos, e assim existiria a necessidade da correção por não-arbitragem apenas nestas duas maturidades para a curva de Cupom Cambial, enquanto que na curva de Eurodólar não podemos rejeitar que todos os fatores de correção são estatisticamente iguais a zero pelos intervalos de credibilidade estimados.

Estes resultados confirmam que a maior liquidez do mercado de Eurodólar já assegura que os dados observados na curva de juros estão livres de oportunidades sistemáticas de arbitragem, enquanto que na curva de cupom cambial estas possibilidades ainda podem estar presentes. Estes resultados são consistentes com os resultados obtidos pela análise de previsão, que mostra que para a curva de Eurodólar a correção de não-arbitragem não altera de forma significativa os resultados de previsão.

Note que estes resultados são no entanto condicionais a estrutura assumida para estes procedimento de verificação de não-arbitragem, que assume que o modelo adequado para a modelagem livre de arbitragem da estrutura a termo de taxas de juros é dado pelo modelo de [Christensen *et al.*, 2008], e assim os resultados obtidos são condicionais a este modelo. Testes com outras forma de correção por não-arbitragem poderia indicar resultados diferentes.

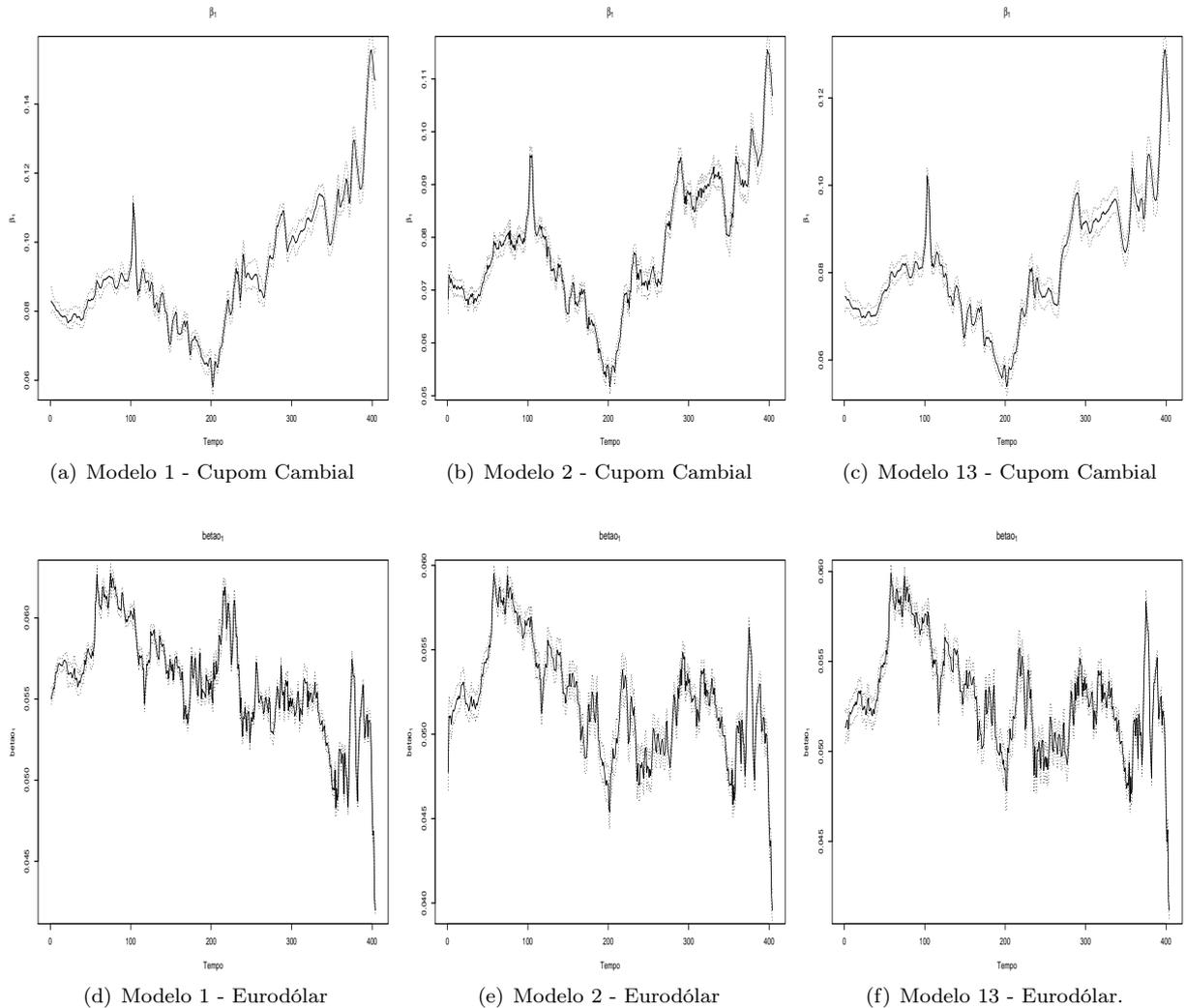
9.4. Fatores Latentes Estimados. Para ilustrar brevemente algumas das características dos modelos estimados, mostramos graficamente algumas comparações entre resultados de distintas especificações. Como existem vários fatores e modelos distintos, apresentamos apenas alguns resultados selecionados¹⁰. Todas as figuras contém a média e os percentis de 2.5% e 97.5% das distribuições posteriores de cada fator latente, representando um intervalo de credibilidade de 95%.

A figura 9.2 mostra a evolução do fator nível (β_1) para os modelos 1 (modelo de curvas independentes), o modelo 2 (modelo de fatores latentes generalizado) e o modelo 13 (modelo de não-arbitragem com Bayesian Shrinkage), para as curvas de Cupom Cambial e Eurodólar. Como é possível notar os resultados são bastante semelhantes entre os diversos modelos, o que está em linha com a estimação deste fator através de distintos modelos para a estrutura a termo (e.g. [Almeida, 2005]), mostrando que os resultados são semelhantes para modelos utilizando ou não a estrutura de não-arbitragem e diferentes especificações.

Os resultados para o fator de nível estimados pela segunda classe de modelos baseados na estrutura de fator global estão colocados na figura 9.3, que mostra os componentes de nível e inclinação estimados. As sub-figuras a) e d) mostram a estimação dos fatores globais de nível e

¹⁰Os demais resultados são disponíveis com os autores.

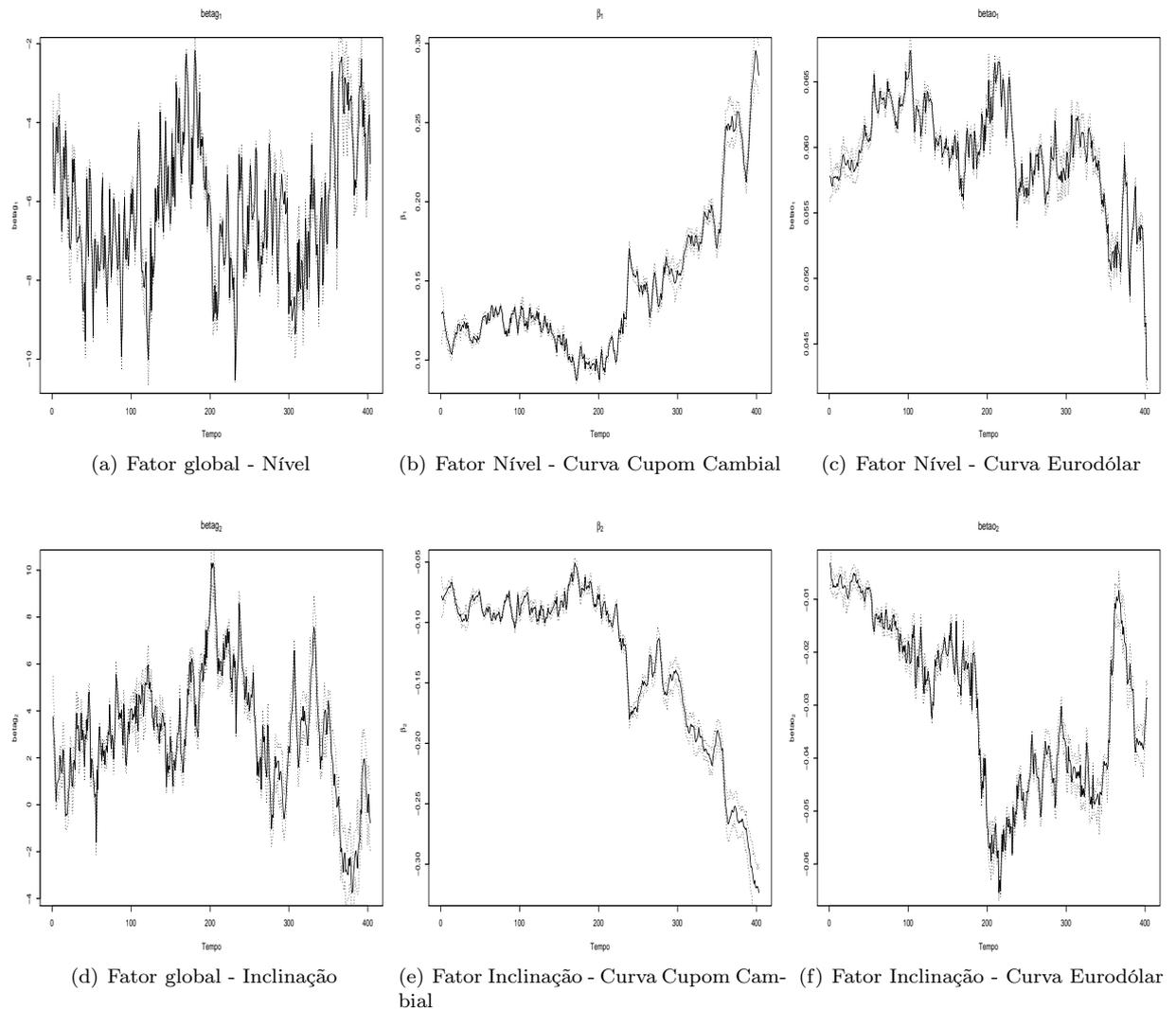
FIGURA 9.2. Fator Nível



inclinação, e as demais sub-figuras mostram as transformações obtidas para a obtenção das curvas de cada mercado através das equações 4.10e 4.11. É possível observar que os fatores globais são mais semelhantes aos fatores obtidos para a curva de Eurodólar, mas também é importante notar que os componentes idiossincráticos são importantes para todas as curvas. É possível notar que os os fatores locais obtidos pelo modelo de fator global generalizado são bastante semelhantes aos obtidos pelos demais modelos estimados, mostrando a consistência na estimação de todos os modelos propostos, e também indicando que a metodologia Bayesiana proposta não sofre com os problemas de identificação. Um problema de identificação seria graficamente evidente se tivéssemos realizações bastante distintas do mesmo fator com poder de ajuste semelhante, o que não acontece com os modelos estimados, já que em todos os modelos os fatores estimados são semelhantes.

A importância de tornar os parâmetros de decaimento τ_1^i e τ_2^i variantes no tempo pode ser observada na Figura 9.4, que mostra a evolução dinâmicas destes parâmetros para as duas curvas de juros modeladas, através da estimação do modelo 2. É possível notar que existe uma significativa variação temporal nestes parâmetros, em especial para o parâmetro τ_1 nas duas curvas, embora o parâmetro τ_2 tenha um comportamento mais ruidoso e com um intervalo de variação menor. Este padrão de variação mostra que esta modificação possibilita aos modelos estimados uma maior

FIGURA 9.3. Modelo de Fatores Globais

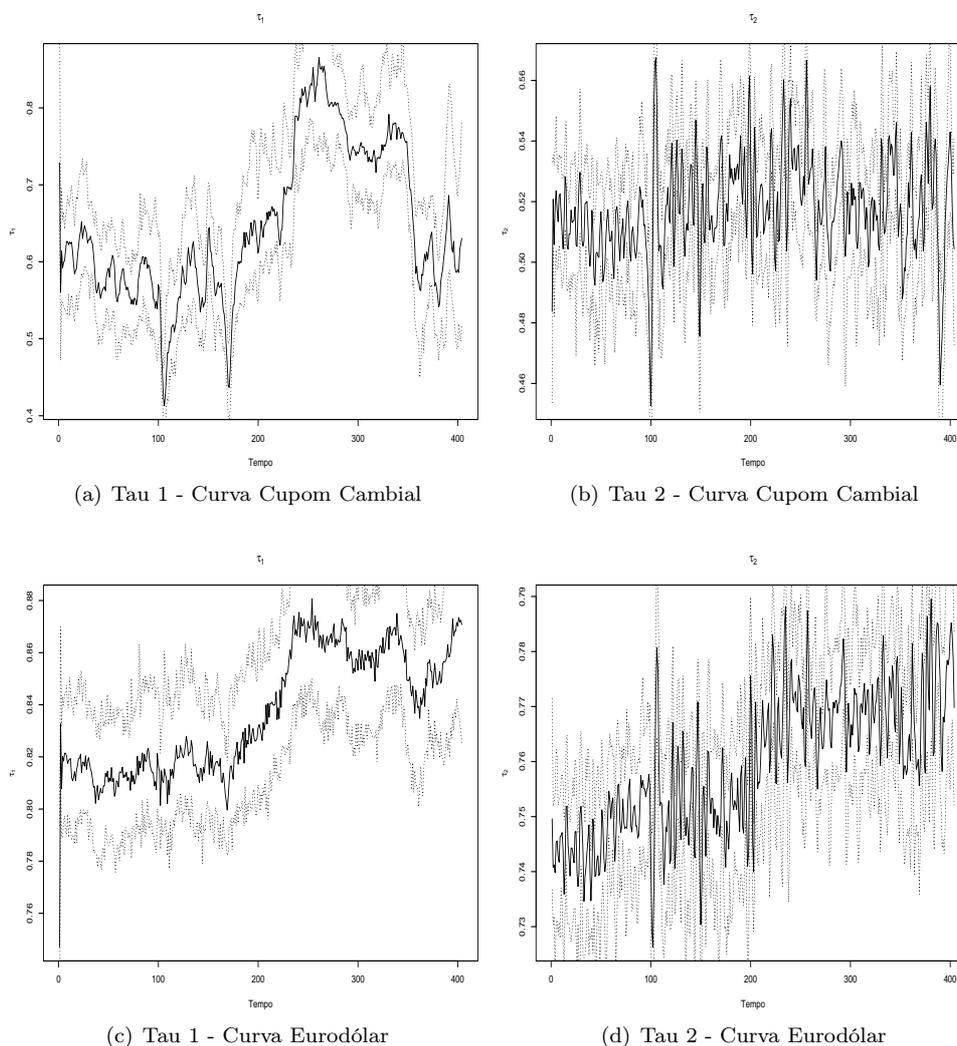


adaptação às mudanças na estrutura a termo de taxas de juros observadas na figura 9.1, e também evita a necessidade de uma especificação ad hoc dos parâmetros de decaimento como utilizada nos artigos de [Diebold & Li, 2006, Diebold *et al.*, 2008].

A validade da utilização de fatores de volatilidade estocástica pode ser visualizado na figura 9.5, que mostra a evolução destes dois fatores estimados através do modelo 2. A dinâmica destes dois fatores é consistente com o padrão de volatilidade observado nas curvas de juros (figura 9.1), acompanhando os períodos de aumento e redução na volatilidade nestas duas curvas, e também mostra que estes fatores latentes adicionais são importantes para identificar corretamente a variação nos demais fatores latentes do modelo. Em todos os modelos com a presença de volatilidade estocástica é possível notar este mesmo comportamento.

A Figura 9.6 mostra alguns exemplos de previsões obtidas pelos modelos propostos. As sub-figuras a) e b) mostram uma comparação das previsões um dia a frente para as curvas de Cupom Cambial e Eurodólar realizadas por todos os modelos, obtidas como médias a posteriori das previsões um passo a frente de cada modelo. A sub-figura c) mostra um exemplo de construção do intervalo de credibilidade de 95% para a previsões um dia a frente para um determinado dia

FIGURA 9.4. Parâmetros de Decaimento



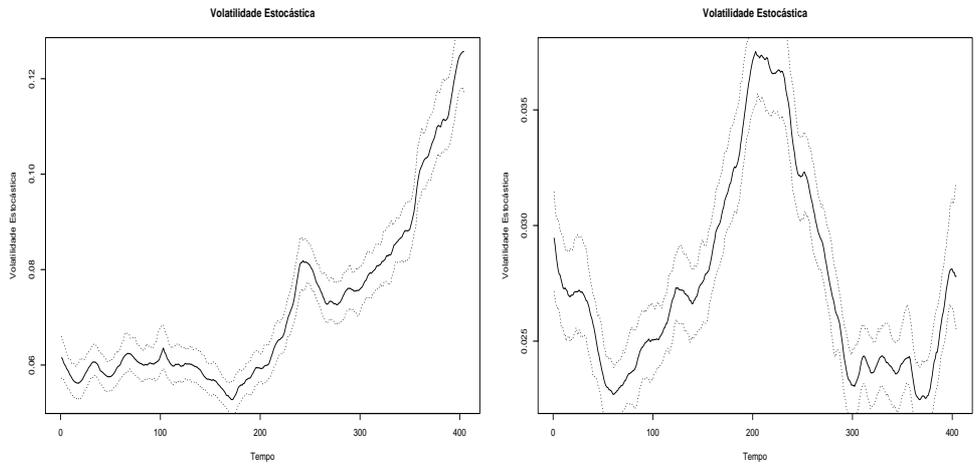
da curva de Eurodólar, neste caso utilizando o modelo 2 de fatores latentes generalizado, e finalmente a sub-figura d) mostra uma comparação entre a previsões utilizando o modelo de 5 fatores (modelo 12, linha contínua) e o modelo equivalente com a correção para não-arbitragem (modelo 13, linha tracejada) para a curva de Cupom Cambial, mostrando que os efeitos da correção por não-arbitragem tem magnitude reduzida, consistente com o resultado geral mostrado na Tabela 5.

Em todos estes exemplos utilizamos diretamente uma propriedade derivada do procedimento de estimação por Markov Chain Monte Carlo, que é a possibilidade de construir intervalos de credibilidade exatos em amostras finitas para os fatores latentes e para as previsões dos modelos. Note que nos procedimentos originais das estimações dos modelos de [Diebold & Li, 2006, Diebold *et al.*, 2008] os intervalos de confiança são construídos sem levar em conta a estimação em dois estágios realizada, e assim só possuem validade assintótica e podem ser bastante viesados em amostras finitas.

10. CONCLUSÕES

Neste trabalho foram propostas uma série de inovações em relação aos procedimentos normalmente utilizados na estimação de modelos para a estrutura a termo de taxas de juros, em especial os

FIGURA 9.5. Volatilidade Estocástica



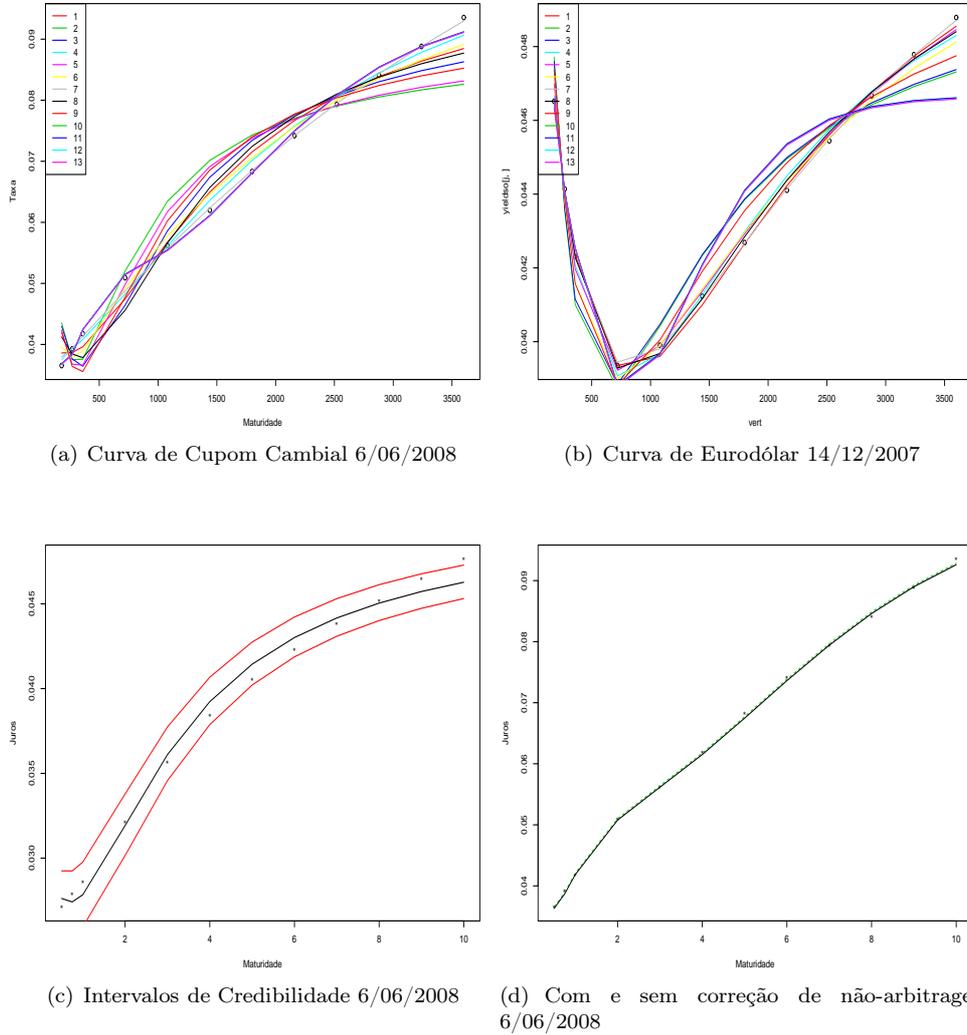
(a) Volatilidade Estocástica - Curva Cupom Cambial (b) Volatilidade Estocástica - Curva Eurodólar

modelos baseados nas especificações de [Diebold & Li, 2006, Diebold *et al.*, 2008, Christensen *et al.*, 2008]. Estas inovações permitem superar diversas limitações e restrições utilizadas nestes modelos, como a escolha da forma funcional, limitada a versões restritas com apenas fatores de nível e inclinação como o modelo adotado em [Diebold *et al.*, 2008], ou então o uso de parâmetros de decaimento fixos e escolhidos de forma ad hoc como na estimação utilizada em [Diebold & Li, 2006]. Os resultados obtidos mostram que existe evidência clara que não somente os fatores latentes evoluem no tempo, mas outros fatores como os parâmetros de decaimento e volatilidade devem ser tratados como fatores latentes adicionais, permitindo procedimentos de ajuste e previsão mais precisos para a estrutura a termo de taxas de juros, especialmente curvas de juros de países emergentes que são caracterizadas por um formato mais rico e com mudanças mais frequentes neste formato.

Os procedimentos de estimação baseados em inferência Bayesiana utilizando algoritmos de Markov Chain Monte Carlo permitem tratar dos problemas que normalmente afetam a estimação de modelos de fatores latentes utilizados em modelagem de taxas de juros, como a existência de máximos locais e os problemas de identificação. A estimação por MCMC não utiliza maximização numérica, e a estrutura de informação a priori e a formulação hierárquica permite contornar os problemas de identificação existentes na estimação de modelos de estrutura a termo de taxas de juros. Esta mesma estrutura de estimação permite reduzir a dimensionalidade do modelo através do uso de Bayesian Shrinkage, um procedimento bastante efetivo como mostrado pelo uso do critério de informação DIC na comparação entre modelos, e assim a estimação destes modelos não necessita de restrições ad hoc como a exclusão de fatores latentes ou a fixação de parâmetros. Os procedimentos de Bayesian Shrinkage propostos são bastante efetivos na redução da dimensionalidade e complexidade dos modelos propostos, um problema especialmente importante no contexto de modelagem conjunta de mais de um mercado.

A inferência Bayesiana é particularmente útil para tratar os problemas relacionados a complexidade dos modelos de estrutura a termo de taxas de juros, caracterizados por estruturas não-lineares e de difícil estimação pelas metodologias clássicas como a estimação por verossimilhança através do filtro de Kalman. O procedimento permite a construção de intervalos de credibilidade exatos, e não é necessária a estimação em vários estágios. O procedimento de estimação por MCMC é interessante já que toda a informação existente na amostra é utilizada diretamente na estimação, já que a estrutura hierárquica em espaço de estados utiliza toda a informação cross-section e no tempo. A estimação Bayesiana permite a estimação de modelos mais complexos e flexíveis para a estrutura a termo de taxas de juros, possibilitando não somente um melhor ajuste mas também

FIGURA 9.6. Previsões



o uso das correções para não-arbitragem que exigem uma estrutura mais complexa de fatores latentes, como mostram [Filipovic, 1999], [Björk & Christensen, 1999] e [Christensen *et al.*, 2008]. Os resultados mostram que utilizando os mecanismos de estimação propostos é possível unir flexibilidade na estimação com consistência com não-arbitragem, possibilitando generalizar estas formulações livres de arbitragem para o ajuste simultâneo de múltiplas curvas de juros.

Esta metodologia de estimação permite obter a distribuição posterior de todos os componentes não-observados, parâmetros e fatores latentes, e estas distribuições podem ser utilizada para verificar outras características importantes, como por exemplo a validade da correção de não-arbitragem através da distribuição posterior do fator de correção de não-arbitragem. Note que este parâmetro é uma função não-linear dos parâmetros de decaimento, e assim sua distribuição não é uma distribuição padrão, e desta forma o uso de procedimentos clássicos de inferência não é trivial, enquanto que na estimação Bayesiana esta informação é um sub-produto padrão do procedimento de estimação.

Os resultados obtidos na aplicação empírica com a modelagem conjunta da curvas de Cupom Cambial e Eurodólares são bastante interessantes. Os resultados mostram que as inovações propostas, como o uso de fatores latentes adicionais para a volatilidade condicional e os parâmetros de decaimento, são efetivas no ajuste e previsão da estrutura a termo destes dois mercados, caracterizados por dinâmicas ricas no formato das curvas. Outro resultado interessante é que a estrutura de interdependência adotada mostra que existe ganho no uso de informação da curva de Eurodólares no ajuste da curva de Cupom Cambial, mas o contrário não é tão importante, e este resultado vai de encontro com o tamanho e a importância relativa destes dois mercados. Estas evidências são confirmadas pela análise preditiva realizada, que confirmam a validade das especificações propostas. Outro ponto interessante é confirmar que a maior liquidez do mercado de Eurodólares impede a existência de oportunidades sistemáticas de arbitragem, o que não ocorre para algumas maturidades no mercado de Cupom Cambial.

REFERÊNCIAS

- [Aldrich, 2002] Aldrich, J. 2002. How Likelihood and Identification went Bayesian. *International Statistical Review*, **70**, 79–89.
- [Almeida, 2005] Almeida, C. I. R. 2005. A Note on the Relation Between Principal Components and Dynamic Factors in Affine Term Structure Models. *Revista de Econometria*, **25(1)**, 89–114.
- [Almeida & Vicente, 2008] Almeida, C. I. R., & Vicente, J. V. M. 2008. The Role of No-arbitrage on Forecasting: Lessons from a Parametric Term Structure Model. *Journal of Banking and Finance*, **32**, 2695–2705.
- [Banbura *et al.*, n.d.] Banbura, M., Giannone, D., & Reichlin, L. *Large Bayesian VARs*. European Central Bank Working Paper.
- [Bauwens *et al.*, 1999] Bauwens, L., Lubrano, M., & Richard, J-F. 1999. *Bayesian Inference in Dynamic Econometric Models*. Cambridge University Press.
- [Bernardo & Smith, 1994] Bernardo, J., & Smith, A. 1994. *Bayesian Theory*. Wiley.
- [Björk & Christensen, 1999] Björk, T., & Christensen, B. J. 1999. Interest Rate Dynamics and Consistent Forward Rate Dynamics. *Mathematical Finance*, **9**, 323–348.
- [Brigo & Mercurio, 2006] Brigo, D., & Mercurio, F. 2006. *Interest Rates Models - Theory and Practice (2nd Edition)*. Springer.
- [Burghardt, 2003] Burghardt, G. 2003. *The Eurodollar futures and Options Handbook*. McGrawHill.
- [Chan *et al.*, 1992] Chan, K. G., Karolyi, G., Longstaff, F., & Sanders, A. 1992. An Empirical Comparison of Alternative Models of Short Term Interest Rate. *Journal of Finance*, **47**, 1209–1297.
- [Christensen *et al.*, 2007] Christensen, J. H., Diebold, F.X., & Rudebusch, G.D. 2007. *The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models*. NBER Working Paper No. 13611.
- [Christensen *et al.*, 2008] Christensen, J. H., Diebold, F. X., & Rudebusch, G. D. 2008. An Arbitrage-Free Generalized Nelson-Siegel Term Structure Model. *Econometrics Journal*, **forthcoming**.
- [Claeskens & Hjort, 2008] Claeskens, C., & Hjort, N, L. 2008. *Model Selection and Model Averaging*. Cambridge University Press.
- [Cogley & Sargent, 2001] Cogley, T., & Sargent, T. 2001. Evolving Post World War II. U.S. Inflation Dynamics. *NBER Macroeconomics Annual*, **16**, 331–373.
- [Cox *et al.*, 1985] Cox, J. C., Ingersoll, J. E., & Ross, S. A. 1985. A Theory of the Term Structure of Interest Rates. *Econometrica*, **53**, 385–408.
- [Dai & Singleton, 2000] Dai, Q., & Singleton, K. 2000. Specification analysis of affine term structure models. *Journal of Finance*, **55**, 1943–1978.
- [Delbaen & Schachermayer, 1994] Delbaen, F., & Schachermayer, W. 1994. A General Version of The Fundamental Theory of Asset Pricing. *Mathematische Annalen*, **300**, 463–520.
- [Diebold *et al.*, 2005] Diebold, F. X., Piazzes, M., & Rudebusch, G. 2005. "Modeling Bond Yields in Finance and Macroeconomics. *American Economic Review*, **95(2)**, 415–420.
- [Diebold & Li, 2006] Diebold, F.X., & Li, C. 2006. Forecasting the Term Structure of Government Bond Yields. *Journal Of Econometrics*, **130**, 337–364.
- [Diebold *et al.*, 2008] Diebold, F.X., Li, C., & Yue, V. 2008. Global Yield Curve Dynamics and Interactions: A Generalized Nelson-Siegel Approach. *Journal of Econometrics*, **146**, 351–363.
- [Doan *et al.*, 1984] Doan, T., Litterman, R., & Sims, C. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**, 1–100.
- [Duffe, 2002] Duffe, G. 2002. Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance*, **57**, 405–443.
- [Duffie & Kan, 1996] Duffie, D., & Kan, R. 1996. A yield-factor model of interest rates. *Mathematical Finance*, **6**, 379–406.
- [Filipovic, 1999] Filipovic, D. 1999. A Note on the Nelson-Siegel Family. *Mathematical Finance*, **9(4)**, 349–359.
- [Filipovic, 2001] Filipovic, D. 2001. *Consistency Problems for Heath-Jarrow-Morton Interest Rate Models*. Springer-Verlag.

- [Florens *et al.*, 1990] Florens, J. P., Mouchard, M., & Rolin, J-M. 1990. *Elements of Bayesian Statistics*. CRC.
- [Gamerman & Lopes, 2006] Gamerman, D., & Lopes, H. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition*. Chapman & Hall/CRC.
- [Harrison & Kreps, 1979] Harrison, J M, & Kreps, D. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, **20**, 381–408.
- [Harrison & Pliska, 1981] Harrison, J M, & Pliska, S. 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, **11**, 215–260.
- [Heath *et al.*, 1992] Heath, David, Jarrow, Robert, & Morton, Andrew. 1992. Bond Pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, **60**(1).
- [Hyndman & Koehler, 2006] Hyndman, R. J., & Koehler, A. B. 2006. Another Look at measures of forecast accuracy. *International Journal of Forecasting*, **22**, 679–688.
- [Kadane, 1974] Kadane, J. B. 1974. *Bayesian Analysis in Econometrics and Statistics*. North-Holland. Chap. The role of identification in Bayesian Theory, pages 175–191.
- [Kadiyala & Karlsson, 2007] Kadiyala, K. R., & Karlsson, S. 2007. Forecasting with generalized bayesian vector auto regressions. *Journal of Forecasting*, **12**, 365 – 378.
- [Kim & Orphanides, 2005] Kim, D. H., & Orphanides, A. 2005. *Term Structure Estimation with Survey Data on Interest Rate Forecasts*. Finance and Economics Discussion Series, 2005-08, Board of Directors of Federal Reserve System.
- [Koop, 2003] Koop, G. 2003. *Bayesian Econometrics*. Wiley.
- [Laurini & Hotta, 2008] Laurini, M. P., & Hotta, L. K. 2008. Bayesian Extensions to Diebold-Li Term Structure Model. In: *Forecasting in Rio*.
- [Litterman & Scheinkman, 1991] Litterman, R., & Scheinkman, J. 1991. Common Factors Affecting Bond Returns. *Journal of Fixed Income*, **1**, 54–61.
- [Lund & Andersen, 1997] Lund, J., & Andersen, T. 1997. Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate. *Journal Of Econometrics*, **77**, 343–377.
- [Morita & Bueno, 2008] Morita, R.H., & Bueno, R.D.S. 2008. Investment Grade Countries Yield Curve Dynamics. In: *63rd. European Meeting of the Econometric Society, 2008, Milão. Annals of the 63rd. European Meeting of the Econometric Society*.
- [Neal, 2003] Neal, R. 2003. Slice Sampling (with discussions). *Annals of Statistics*, **31**, 705–767.
- [Nelson & Siegel, 1987] Nelson, C. R., & Siegel, A. F. 1987. Parsimonious Modelling of Yield Curves. *Journal of Business*, **60**(4), 473–489.
- [Ntzoufras, 2009] Ntzoufras, I. 2009. *Bayesian Modeling Using WinBUGS*. Wiley.
- [Park & Casella, 2008] Park, T., & Casella, G. 2008. The Bayesian Lasso. *Journal of the American Statistical Association*, **103**, 681–686.
- [Pereira, 2009] Pereira, F. T. G. 2009. *Curva a termo para o risco de convertibilidade: Uma abordagem utilizando o diferencial de juros*. working paper não publicado.
- [Pinheiro *et al.*, 2007] Pinheiro, F., Almeida, C. I., & Vicente, J. 2007. Um modelo de Fatores Latentes com Variáveis Macroeconômicas para a Curva de Cupom Cambial. *Revista Brasileira de Finanças*, **5**(1), 79–92.
- [Poirier, 1998] Poirier, D. J. 1998. Revising Beliefs in Nonidentified Models. *Econometric Theory*, **14**, 483–509.
- [Robert & Casella, 2005] Robert, C. P., & Casella, G. 2005. *Monte Carlo Statistical Methods*. Springer.
- [Robertson, 1999] Robertson, J. C. and Tallman, E. W. 1999. Vector autoregressions: forecasting and reality. *Economic Review*, **Q1**, 4–18.
- [Rothemberg, 1971] Rothemberg, T. 1971. Identification in Parametric Models. *Econometrica*, **39**, 577–591.
- [Sims, 2001] Sims, C. 2001. Comment on Sargent and Cogley’s Evolving Post World War II U.S. Inflation Dynamics. *NBER Macroeconomics Annual*, **16**, 373–379.
- [Singleton, 2006] Singleton, K. J. 2006. *Empirical Dynamic Asset Pricing*. Princeton University Press.
- [Spiegelhalter *et al.*, 2002] Spiegelhalter, D., Best, N. G., Carlin, B. P, & van der Linde A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **64**(4), 583–639.
- [Svensson, 1994] Svensson, L. E. O. 1994. Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994. *NBER Working Paper*.
- [Tibshirani, 1996] Tibshirani, R. 1996. Regression Shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.
- [Tierney, 1994] Tierney, L. 1994. Markov Chains for Exploring Posterior Distributions (with Discussion). *Annals of Statistics*, **22**, 1701–1786.
- [Vasicek, 1977] Vasicek, O. 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177–88.